Predicting the outcome of physiotherapy in people with painful partial-thickness rotator cuff tears

Cordula Braun

A thesis submitted in partial fulfilment of the requirements
of Teesside University for the award of the degree of
Doctor of Philosophy (PhD)

Health and Social Care Institute
School of Health and Social Care
Teesside University
Middlesbrough, UK

July 2016

Predicting the outcome of physiotherapy in adults with painful partial-thickness rotator cuff tears

Cordula Braun

Supervisors:

Dr Nigel Hanchard (Director of Studies)

Dr Helen Handoll

Prof Alan Batterham

Local advisor (Germany):

Dr med Andreas Betthäuser

Note on final thesis copy

The thesis was initially presented in two volumes, Volume 1 (Main Thesis) and Volume 2 (Appendices). These have been combined, so that the complete thesis is now presented in one volume.

Due to copyright issues, this thesis copy does not include two appendices 3.1 and 4.1.

Declaration

I hereby declare that the work presented in this thesis represents entirely my own work, and that no part of it has been submitted in support of an application for any other qualification, degree or award of this or any other university or institute of learning.

Abstract

Rotator cuff disorders encompass a range of impairments from tendinopathy to partial- or full-thickness rotator cuff tears, and represent the largest subgroup of shoulder pain. Rotator cuff tears, most of which are atraumatic, are common in adults with shoulder pain and are strongly associated with increasing age. Conservative treatment including physiotherapy is the first-line treatment, but some patients do not respond, and ultimately require surgery. Early predictions of response could allow individuals' care pathways to be optimised, preventing unnecessary delays and suffering and benefiting patients and healthcare providers alike.

My primary aim was to develop a prognostic model for the outcome of physiotherapy in adults with painful atraumatic partial-thickness tears (PTTs) of the rotator cuff. This was addressed by a prospective prognostic model study. The study was underpinned by a systematic review of prognostic models in adults undergoing physiotherapy for painful rotator cuff disorders and was further informed and complemented by the following work: the development and validation of the physiotherapy protocol for the prognostic study; the identification, selection and definition of the candidate prognostic factors for the prognostic study; the estimation of the Minimal Important Difference (MID) of the study's primary outcome measure (the Western Ontario Rotator Cuff Index, WORC); and an exploratory responder analysis of the WORC outcome scores. The prognostic systematic review, prognostic study, MID analysis and responder analysis are original contributions to knowledge.

The prognostic systematic review revealed important methodological deficiencies in the five included studies, and no clinically usable model. No study addressed a distinct PTT population. The process of identifying factors for my own prognostic model study revealed a lack of knowledge about the prognostic relevance of factors. All of the candidate models I explored in my prognostic study (n sample = 65, n analysed = 61) had low performance and precision. The estimated MID of the WORC was -300. The responder analysis resulted in different proportions of responders to treatment depending on the responder definition.

My results highlight the difficulties involved in predicting outcomes in the field of shoulder pain and rotator cuff disorders, and the need for methodologically sound prognosis research.

Acknowledgements

I would like to express my sincere thanks to my fantastic supervisory team: first of all, to Dr Nigel Hanchard for giving me the opportunity to embark on this PhD journey by agreeing to be my Director of Studies, and for fulfilling this role so exceptionally by providing invaluable guidance, assistance and encouragement throughout, from the very initial idea through to the completion of this thesis; to Dr Helen Handoll for providing support on many aspects of my PhD, and in particular for always being available to discuss things through and provide advice and help, and for travelling all the way to Teesside for supervision whenever I was over from Germany; and to Prof Alan Batterham for providing statistical support on the analysis of the prognostic study and the MID and responder analyses, and in particular for advising me on the novel analysis approaches which I applied.

I would like to sincerely thank Dr Andreas Betthäuser for his invaluable contributions to my prognostic study as the person who recruited, assessed and monitored all participants, and for acting as an external advisor to my PhD.

I would like to thank all who further contributed to my PhD and the thesis in various ways, in particular: all patients who participated in the prognostic study; all collaborating physiotherapists; Dr Robert Wolff for letting me use the PROBAST tool in the prognostic systematic review, and for providing support on its use; Julie Hogg and Iain Baird for providing advice on the development of the search strategy for the prognostic systematic review; everyone at the Health and Social Care Institute for their continuous help and administrative assistance; the management of hochschule 21, Buxtehude, Germany, for their generous support of my PhD, and my colleagues for all their help; and Jürgen Sandfort for assisting with the design of the images in this thesis.

I would like to thank my family and all of my friends for accompanying me on this journey, and for providing support and encouragement, but, most importantly, for always being there for me.

Lastly, I would like to express my deepest thanks to my parents for their priceless contributions in so many ways. Without them, I would not have been able to embark on a PhD, and would not be where I am today. I am very sad that my mother isn't with us anymore.

Scientific Output

Publications

1. Braun C., Bularczyk M., Heintsch J., Hanchard N. C. A. (2013). Manual therapy and exercises for shoulder impingement revisited. *Phys Ther Rev*, *18*(4), 263–84. doi:10.1179/108331913X13709388114510.

2. Braun C., Hanchard N. C., Batterham A. M., Handoll H. H., Betthäuser A. (2015). Prognostic Models in Adults Undergoing Physical Therapy for Rotator Cuff Disorders: Systematic Review *Phys Ther*, Dec 4 [Epub ahead of print][1]. doi:10.2522/ptj.20150475.

Presentations

1. Oral presentation: *"Prognose und Prognoseforschung in der Physiotherapie" ["Prognosis and prognosis research in physiotherapy"].* Bundeskongress Physiotherapie [German national conference for physiotherapy], Berlin-Brandenburg, Van der Valk Hotel, 19 Sep 2014.

included background work on prognosis and prognosis research and an outline of the prognostic model study.

2. Oral presentation (invited): *"Evidenzbasierte Therapie des subakromialen Schmerzes: physiotherapeutische Maßnahmen" [Evidence-based treatment of subacromial pain: physiotherapy interventions].*

A) Schulternetzwerk Deutschland [Shoulder Network Germany], Netzwerktreffen [network meeting], Hamburg, Hochschule Fresenius, 24 Feb 2016; and

B) Schulternetzwerk Hamburg [Shoulder Network Hamburg], IV. Qualitätszirkel [quality circle], Hamburg, Evangelisches Krankenhaus Alsterdorf, 17 Feb 2016.

included a summary of the intervention systematic review and a brief summary of the prognostic systematic review.

[1] The article is scheduled for the July 2016 issue of *Physical Therapy*.

Preface

"It is difficult to make predictions,
particularly about the future."

(attributed to Mark Twain[2])

[2] See http://quoteinvestigator.com/2013/10/20/no-predict/ [Last accessed 22 June 2016]

Table of contents

Copyright note

List of Tables

List of Figures

List of Textboxes

Abbreviations and acronyms

The following list describes abbreviations and acronyms that are used repeatedly in the main thesis text. Abbreviations and acronyms which are exclusively used in tables are not included; they are described in the respective table keys. Standard abbreviations (e.g. yr, wk) are also not included.

AC joint	Acromioclavicular joint
ADJ, $_{ADJ}$	Adjusted (in this thesis: for regression to the mean, RTM)
AIC	Akaike's Information Criterion
AIC_C	Akaike's Information Criterion, small sample variant
□AIC_C	AIC_C difference
AIC_{MIN}	Smallest AIC (AIC_c) value
AMSTAR	Assessing the Methodology of Systematic Reviews (checklist)
CI	Confidence interval
DEGUM	Deutsche Gesellschaft für Ultraschall in der Medizin [German Society for Ultrasound in Medicine]
FTT	Full-thickness tear (rotator cuff)
GKV	Gesetzliche Krankenversicherung [statutory health insurance]
GPC	Global Perceived Change
HRQoL	Health-related quality of life
ICD, ICD-10	International Classification of Diseases, 10th revision
ICTRP	International Clinical Trials Registry Platform
KL information	Kullback-Leibler information
LHB	Long head of biceps
MCID	Minimal Clinically Important Difference
MID	Minimal Important Difference
MRA	Magnetic resonance arthrography
MRI	Magnetic resonance imaging
NRS	Numerical rating scale
NSAID	Nonsteroidal anti-inflammatory drug
PCS	Pain Catastrophizing Scale
PEDro	Physiotherapy Evidence Database
PROBAST	Prediction Risk of Bias and Applicability Tool
PROGRESS	PROGnosis RESearch Strategy Partnership
PROM	Patient-reported outcome measure
PTT	Partial-thickness tear (rotator cuff)

R^2	Coefficient of (multiple) determination
RC-QUOL	Rotator Cuff Quality of Life tool
RCT	Randomised controlled trial
ROM	Range of movement
RTM	Regression to the mean
SD	Standard deviation
SD_{IR}	Standard deviation of individual responses
SEE	Standard error of the estimate
SEM	Standard error of measurement
US	Ultrasonography
VAS	Visual analogue scale
WORC	Western Ontario Rotator Cuff index
WORC_1	Baseline WORC
WORC_2	Follow-up WORC
WORC_change	Change of WORC score from baseline to follow-up

PART ONE:

INTRODUCTION AND BACKGROUND

CHAPTER 1
General introduction, aims, content and structure of the thesis

1.1 TOPIC

The shoulder joint is the primary link between the upper extremity and upper body. Its exceptional mobility, the greatest of any human joint (Kapandji 1992 p. 2), facilitates positioning of the upper extremity and, ultimately, the hand in space. Painless and unimpaired shoulder functioning is indispensible for many activities of daily living, such as washing, putting on a coat or combing one's hair, and an important prerequisite for participation in occupational and recreational activities (*Figure 1.1*).

Figure 1.1: Functional relevance of the shoulder
(Silhouettes modified from cliparts available at www.all-silhouettes.com*)*

The particular anatomical composition of the glenohumeral joint, with its large humeral articular surface in relation to the significantly smaller surface of the glenoid socket, makes it highly dependent on its non-osseous components, specifically its shoulder muscles, which thus play a key role in normal kinematics (Hess 2000).

Shoulder pain is common, with an estimated point prevalence of 7-26%, one-year prevalence of 5-47% and lifetime prevalence of 7-67% in the general population, as most recently reviewed by (Luime et al. 2004). Painful shoulder complaints are the second to third commonest type of musculoskeletal pain seen in general medical practice (Kooijman et al. 2013, Urwin et al. 1998). They can be persistent and can lead to increased use of healthcare resources and prolonged sick leave, consistently placing a cost burden on the individual and society (Leps et al. 2012, Paloneva et al. 2013, van der Windt et al. 1996, Virta et al. 2012). In the context of the International Classification of Diseases, 10th revision (ICD-10) (WHO 2016), shoulder lesions (ICD-10 code M75) consistently rank among the 10 diseases that cause most time off work in Germany (DAK Forschung 2015, AOK 2008)[3]. No current data are available on the precise (direct and indirect) costs of shoulder lesions in Germany, but an economic analysis of data from 2002 estimated the overall annual costs at 2.1 billion euros (approximately 6% of the overall costs of all musculoskeletal diseases). Of these, approximately 60% (1.25 billion euros) represented indirect costs related to lost working days due to inability to work (Leps et al. 2012).

Shoulder pain can be due to various problems and pathologies, but most cases (29% to 85%) involve the subacromial-subdeltoid bursa and the rotator cuff (Cadogan et al. 2011, Dorrestijn et al. 2011, Juel & Natvig 2014, Östör et al. 2005, Tekavec et al. 2012, van der Windt et al. 1995). The programme of research that is presented within this thesis aimed to inform the conservative treatment of adults with shoulder pain in the presence of a specific disorder of the rotator cuff: a painful, partial-thickness rotator cuff tear (PTT).

1.2 AIMS AND CONTENT

This thesis reports a body of research aimed at informing the conservative treatment of people with painful PTTs. The primary aim was to develop a prognostic model for the outcome of a phase of conservative treatment with physiotherapy in adults with painful, atraumatic PTTs.

[3] Throughout this thesis there is a focus on German data (where available) because the prognostic study was conducted in Germany.

□

The research comprises the following components:

1) a systematic review of the evidence for the effectiveness of physiotherapy interventions for patients with impingement-related shoulder pain;
2) a systematic review of the evidence on prognostic models for predicting outcomes in adults undergoing physiotherapy for painful rotator cuff disorders;
3) a prognostic cohort study to develop a prognostic model for predicting the outcome of a phase of conservative treatment with physiotherapy in adults with painful atraumatic PTTs; and
4) an analysis to estimate the Minimal Important Difference (MID) of the Western Ontario Rotator Cuff Index (WORC), the primary outcome measure used in the prognostic study. This analysis is complemented by an exploratory responder analysis.

1.3 STRUCTURE

The thesis is structured into four parts and eight chapters. The structure, content and specific aims are summarised in *Orientation Table Chapter 1* (see next page). As an aid to navigation through the thesis, a similar orientation table is placed on the page before each new chapter (*Orientation Tables Chapters 2 to 8*). As in *Orientation Table Chapter 1*, the chapters to which these respectively pertain are emphasised by bold and unshaded font. The references are provided at the end of each chapter. The thesis is presented in two volumes: *Volume 1* includes *Parts One to Three* (*Chapters 1-8*), and *Volume 2* includes *Part Four* (*Appendices*).

The research is presented in *Part Two*. The prognostic systematic review (*Chapter 3*) is reported first because it contextualises and justifies my prognostic study. The review was formally conducted in 2014 to 15, but builds on systematic literature searches which I conducted at the beginning of my PhD to determine the rationale for my planned study and to inform its design. The two subsequent chapters (*Chapters 4* and *5*) further underpin the prognostic study, which is reported in *Chapter 6*. *Chapter 7*, which reports on the MID and responder analyses, relates to the observed primary outcome of the prognostic study.

Orientation Table Chapter 1

Part	Ch.	Title	Aims
ONE	**1**	**General introduction, aims, content and structure of the thesis**	**1. To provide a general introduction to the topic** **2. To summarise the aims, content and structure of the thesis**
	2	Background	To provide the relevant topical and conceptual background to the programme of research
TWO	3	Prognostic models in adults undergoing physiotherapy for rotator cuff disorders - a systematic review	To establish the state of evidence on prognostic models in adults undergoing physiotherapy for painful rotator cuff disorders
	4	Developing and validating the physiotherapy protocol for the prognostic study	1. To establish the state of evidence on the effectiveness of physiotherapy interventions for adults with painful atraumatic PTTs 2. To develop and validate the physiotherapy treatment protocol
	5	Selecting and defining the candidate prognostic factors for the prognostic study	1. To identify and select the candidate factors for the prognostic model study 2. To define the specific measures for the selected factors
	6	Predicting the outcome of physiotherapy in adults with painful partial-thickness rotator cuff tears (PTTs) – a prognostic model study	To develop a prognostic model for the outcome of a period of conservative treatment with physiotherapy in adult patients with painful atraumatic PTTs
	7	Drawing meaningful magnitude-based inferences from the prognostic study – Minimal Important Difference (MID) and responder analysis	1. To establish an estimate of the MID of the Western Ontario Rotator Cuff Index (WORC), the primary outcome of the prognostic model study 2. To apply the estimated MID to an exploratory responder analysis
THREE	8	Overall summary and conclusions	1. To summarise the research 2. To provide overall conclusions and consider implications
FOUR		Appendices	Appendices to Chapters 3-7

1.4 NOTE ON LANGUAGE OF DOCUMENTATION AND MATERIALS WITHIN THE THESIS

Unless stated otherwise, all German-language documentation and materials are appended in English translation.

1.5 REFERENCES

AOK (2008). (Federation of General Local Health Insurance Funds, AOK-BV). *The 10/20/50 diseases with the longest periods of inability to work of compulsory members of the Local Statutory Health Insurance (AOK) without pensioners*. Available at: http://www.gbe-bund.de (Search by topic: shoulder lesions Expenditures, Costs and Financing The 10/20/50 diseases…*;* document type: table) [Last accessed 03 June 2016].

Cadogan A., Laslett M., Hing W. A., McNair P. J., Coates M. H. (2011). A prospective study of shoulder pain in primary care: prevalence of imaged pathology and response to guided diagnostic blocks. *BMC Musculoskelet Disord*, 12(1), 119. doi:10.1186/1471-2474-12-119.

DAK Forschung (2015). (German Employees' Health Insurance, DAK). *DAK Gesundheitsreport* [*DAK Health Report*]. Available at: https://www.dak.de/dak/download/Vollstaendiger_bundesweiter_Gesundheitsreport_2015-1585948.pdf [Last accessed: 03 June 2016].

Dorrestijn O., Greving K., van der Veen W. J., van der Meer K., Diercks R. L., Winters J. C., Stevens M. (2011). Patients with shoulder complaints in general practice: consumption of medical care. *Rheumatology (Oxford)*, 50(2), 389-95. doi:10.1093/rheumatology/keq333.

Hess S. A. (2000). Functional stability of the glenohumeral joint. *Man Ther*, 5(2), 63-71.

Juel N. G., Natvig B. (2014). Shoulder diagnoses in secondary care, a one year cohort. *BMC Musculoskelet Disord*, 15(1), 89. doi:10.1186/1471-2474-15-89.

Kapandji I. A. (1992). *Funktionelle Anatomie der Gelenke. Band 1: Obere Extremität* [Functional anatomy of the joints. Volume 1: Upper Extremity]. 2nd edn. Stuttgart: Enke.

Kooijman M., Swinkels I., van Dijk C., de Bakker D., Veenhof C. (2013). Patients with shoulder syndromes in general and physiotherapy practice: an observational study. *BMC Musculoskelet Disord*, 14(1), 128. doi:10.1186/1471-2474-14-128.

Leps C., Falz R., Sauer J., Bischoff C., Friedrich R., Schulze A., Thomas M., Busse M. (2012). Epidemiologische und ökonomische Bewertung von Muskel-Skelett-Erkrankungen für die ICD-10 Klasse M75 Schulterläsionen [Epidemiological and economic valuation of musculoskeletal disorders for the ICD-10 class M75 shoulder lesions]. *Klin Sport*, 13(1), 1-6.

Luime J. J., Koes B. W., Hendriksen I. J. M., Burdorf A., Verhagen A. P., Miedema H. S., Verhaar J. A. N. (2004). Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol*, 33(2), 73-81. doi:10.1080/03009740310004667.

Östör A. J. K., Richards C. A., Prevost A. T., Speed C. A., Hazleman B. L. (2005). Diagnosis and relation to general health of shoulder disorders presenting to primary care. *Rheumatology (Oxford)*, 44(6), 800-5. doi:10.1093/rheumatology/keh598.

Paloneva J., Koskela S., Kautiainen H., Vanhala M., Kiviranta I. (2013). Consumption of medical resources and outcome of shoulder disorders in primary health care consulters. *BMC Musculoskelet Disord*, 14(1), 348. doi:10.1186/1471-2474-14-348.

Tekavec E., Jöud A., Rittner, R., Mikoczy Z., Nordander C., Petersson I. F., Englund M. (2012). Population-based consultation patterns in patients with shoulder pain diagnoses. *BMC Musculoskelet Disord*, 13(1), 238. doi:10.1186/1471-2474-13-238.

Urwin M., Symmons D., Allison T., Brammah T., Busby H., Roxby M., Simmons A., Williams G. (1998). Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Ann Rheum Dis*, 57(11), 649-55. doi:10.1136/ard.57.11.649.

van der Windt D. A., Koes B. W., de Jong B. A., Bouter L. M. (1995). Shoulder disorders in general practice: incidence, patient characteristics, and management. *Ann Rheum Dis*, 54(12), 959-64. doi:10.1136/ard.54.12.959.

van der Windt D. A. W. M., Koes B. W., Boeke A. J. P., Devillé W., de Jong B. A., Bouter L. M. (1996). Shoulder disorders in general practice: Prognostic indicators of outcome. *Br J Gen Pract*, 46(410), 519-23.

Virta L., Joranger P., Brox J. I., Eriksson, R. (2012). Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. *BMC Musculoskelet. Disord*, 13(1), 17. doi:10.1186/1471-2474-13-17.

WHO (2016). *International Classification of Diseases (ICD).* Available at http://www.who.int/classifications/icd/en/ [Last accessed 10 June 2016].

Orientation Table Chapter 2

Part	Ch.	Title	Aims
ONE	1	General introduction, aims, content and structure of the thesis	1. To provide a general introduction to the topic 2. To summarise the aims, content and structure of the thesis
	2	**Background**	**To provide the relevant topical and conceptual background to the programme of research**
TWO	3	Prognostic models in adults undergoing physiotherapy for rotator cuff disorders - a systematic review	To establish the state of evidence on prognostic models in adults undergoing physiotherapy for painful rotator cuff disorders
	4	Developing and validating the physiotherapy protocol for the prognostic study	1. To establish the state of evidence on the effectiveness of physiotherapy interventions for adults with painful atraumatic PTTs 2. To develop and validate the physiotherapy treatment protocol
	5	Selecting and defining the candidate prognostic factors for the prognostic study	1. To identify and select the candidate factors for the prognostic model study 2. To define the specific measures for the selected factors
	6	Predicting the outcome of physiotherapy in adults with painful partial-thickness rotator cuff tears (PTTs) – a prognostic model study	To develop a prognostic model for the outcome of a period of conservative treatment with physiotherapy in adult patients with painful atraumatic PTTs
	7	Drawing meaningful magnitude-based inferences from the prognostic study – Minimal Important Difference (MID) and responder analysis	1. To establish an estimate of the MID of the Western Ontario Rotator Cuff Index (WORC), the primary outcome of the prognostic model study 2. To apply the estimated MID to an exploratory responder analysis
THREE	8	Overall summary and conclusions	1. To summarise the research 2. To provide overall conclusions and consider implications
FOUR		Appendices	Appendices to Chapters 3-7

CHAPTER 2
Background

2.1 FUNCTIONAL RELEVANCE OF THE ROTATOR CUFF

The rotator cuff is a deep cuff of four tendons around the shoulder through which the supraspinatus, infraspinatus, teres minor and subscapularis muscles (*Figure 2.1*) attach to the humeral tuberosities. The rotator cuff holds a primary role in the dynamic stabilisation of the glenohumeral joint (Escamilla et al. 2009, Hess 2000, Labriola et al. 2005), mediating important, complex contributions to shoulder movement and providing dynamic glenohumeral compression. By counteracting the superiorly directed forces of shoulder muscles such as deltoid, the rotator cuff prevents superior migration of the humeral head (Escamilla et al. 2009, Magarey & Jones 2003). The rotator cuff, but specifically supraspinatus, has a close positional relationship to other structures of the shoulder such as the subacromial-subdeltoid bursa (*Figure 2.1*) and the acromioclavicular (AC) joint (not shown). The long head of biceps (LHB), despite its proximity, is usually viewed not as part of the rotator cuff but a separate anatomical entity, with distinct pathophysiologies (Ejnisman et al. 2010).

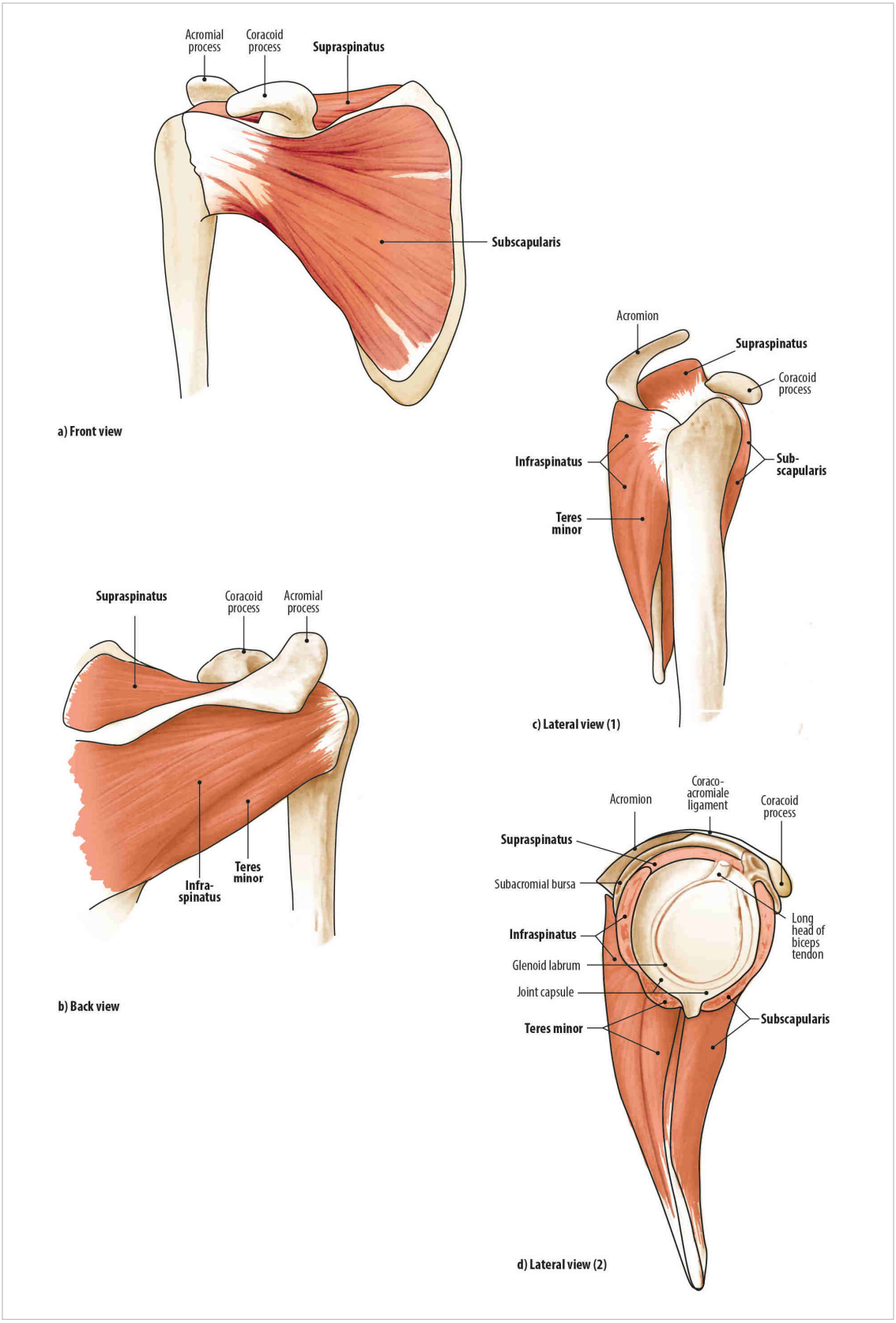


Figure 2.1: Anatomy of the rotator cuff.
(Modified from: Schünke M., Schulte E., Schumacher U. (eds.) (2005). *Allgemeine Anatomie und Bewegungssystem (Prometheus Lernatlas der Anatomie).* Stuttgart: Thieme, pp. 232 & 263)

2.2 PARTIAL-THICKNESS ROTATOR CUFF TEARS

Rotator cuff tears are characterised by either partial or complete discontinuity of the affected tendon (Hedtmann 2009). They are accordingly termed partial-thickness tears (PTTs) or full-thickness tears (FTTs). The supraspinatus is the most frequently affected tendon by far (Matava et al. 2005), and is involved in between 63% (Yamanaka & Matsumoto 1994) and 100% (Maman et al. 2009) of PTTs. It is also usually the first tendon to tear (Beaudreuil et al. 2007, Hedtmann 2009). Supraspinatus PTTs may extend into the infraspinatus, teres minor and rarely also into the subscapularis, but isolated tears of these other tendons are rare, as is the involvement of all (Beaudreuil et al. 2007, Hedtmann 2009)

PTTs are commonly classified by their location into articular-sided, bursal-sided or intratendinous tears (Finnan & Crosby 2010, Smith et al. 2010). Articular-sided tears represent the commonest type. They are twice as common as bursal-sided tears (Hedtmann 2009, Matava et al. 2005, Smith et al. 2010). Other classifications further consider the extent of the tear (Ellman 1990, Habermeyer et al. 2006, Snyder et al. 1991).

2.3 PREVALENCE

The precise prevalence of PTTs is unclear (Finnan & Crosby 2010, Shin 2011) and varies across the literature depending on the type of study (which may involve cadavers, symptomatic patients or asymptomatic participants) and the methods used to determine tears (direct visualisation at dissection or open surgery, arthroscopy, ultrasonography (US) or magnetic resonance imaging (MRI)). Very few imaging studies have investigated the prevalence of PTTs in painful (symptomatic) clinical populations (Cadogan et al. 2011, Reilly et al. 2006, Yamaguchi et al. 2006). Reilly et al. (2006), in a systematic review of prevalence studies, reported a prevalence of PTTs in shoulder pain populations of 7% as determined by US (based on 9 studies, 1038 participants, mean age 50 years), and of 9% as determined by MRI (12 studies, 490 participants, mean age 44 years). Two more recent studies (Cadogan et al. 2011, Yamaguchi et al. 2006) found significantly higher rates of 23% (Cadogan et al. 2011) and 24% (Yamaguchi et al. 2006) as determined by US (in 203 and 588 participants with a mean age of 42 and 63 years, respectively). However, PTTs and FTTs alike are also frequently present in asymptomatic individuals: for PTTs, Reilly et al. (2006) reported a prevalence of 17% as determined by US and of 16% as determined by MRI. To date

(May 2016), no data are publicly available on the prevalence of rotator cuff tears in the German population.

The prevalence of degenerative, atraumatic rotator cuff tears is associated with increasing age (Lashgari & Redziniak 2012, Tashjian 2012), and they are very rare below the age of 40 years (Beaudreuil et al. 2007, Hedtmann 2009). The reported mean age of study populations with painful PTTs ranges from 43 (Gartsman & Milne 1995) to 61 (Yamanaka & Matsumoto 1994). There are no clear data on the gender distribution of painful atraumatic PTTs.

2.4 LINKING PTTs AND SHOULDER PAIN

The fact that PTTs are common in both symptomatic and asymptomatic shoulders suggests that PTTs may be clinically irrelevant unless associated with shoulder pain and disability (Reilly et al. 2006). The precise source of pain in the presence of rotator cuff tears is yet unclear (Khan et al. 2000). It is recognised, though, that the pain cannot be unambiguously attributed to the tendon itself: some studies have found an association between pain and the presence and extent of a (subacromial) “bursal reaction”, evidenced by increased amounts of substance P (a nociceptive neurotransmitter) and histological findings (Gotoh et al. 1998, Ishii et al. 1997). Several studies (Fukuda 2000, Gotoh et al. 1998, Gschwend et al. 1988, Heers et al. 2005) have observed higher levels of pain in people with PPTs compared with FTTs, which suggests that symptomatic PTTs may be more painful than FTTs. It has been supposed that this difference may be related to the amount of bursal tissue or the distribution of substance P (Gotoh et al. 1998). The relationship between the size of PTTs and shoulder pain is also largely unclear. Based on limited evidence there is no association between the size and thickness of PTTs and pain severity (Curry et al. 2015).

Throughout this thesis, the term “painful PTTs” (or “painful rotator cuff tears”) is used to concisely label the population of interest: adults with *shoulder pain in the presence of PTTs*. The uncertainties regarding the link between shoulder pain and PTTs are acknowledged, however.

2.5 CLINICAL PRESENTATION

The clinical presentation of painful PTTs is essentially that of “shoulder impingement” (Bayam et al. 2011, Finnan & Crosby 2010, Fukuda 2003, Hedtmann 2009). The pain

commonly occurs in the anterolateral aspect of the shoulder; it is often perceived as sharp and may be accompanied by a dull pain in the lower arm or forearm (Bayam et al. 2011). The pain is provoked or aggravated with various movements or activities of the arm at or above shoulder level. People with PTTs commonly have nocturnal pain (Finnan & Crosby 2010; Fukuda 2003). While the ability to move the arm may be limited through pain inhibition, both shoulder passive range of movement (ROM) and strength are usually largely preserved (Hedtmann 2009).

2.6 INDIVIDUAL AND SOCIOECONOMIC BURDEN

Rotator cuff tears can significantly affect people's lives by impairing shoulder function, activities and health-related quality of life (HRQoL) (Piitulainen et al. 2012, Ryliskis et al. 2009). A recent qualitative study (Minns Lowe et al. 2014), though limited to people with FTTs, provides an impression of patients' perspectives of living with a painful rotator cuff tear. Analysis of semi-structured interviews with 20 patients revealed that rotator cuff tears can cause severe pain and significantly affect sleep, mobility, strength, activities of daily living, recreational or occupational tasks and emotional wellbeing, as well as imposing a financial burden due to inability to work and private healthcare costs.

No data are publicly available on the burden of costs associated with rotator cuff tears (or any other subgroup of shoulder lesions (ICD-10 code M75)) in the German population (Leps et al. 2012), but German hospital data show a growing trend. The number of admissions for rotator cuff lesions (ICD-10 code M75.1) almost doubled between 2005 (24,500) and 2014 (46,600) (Statistisches Bundesamt 2016). These data may represent the tip of an iceberg, because more than 90% of patients with shoulder complaints are treated in the outpatient setting (Leps et al. 2012).

2.7 AETIOLOGY AND PATHOGENESIS

Over 90% of rotator cuff tears are considered to be caused by degeneration rather than by trauma (Mall et al. 2013, Oh et al. 2007, Petersen & Murphy 2011). The aetiology of the degeneration is probably multifactorial (Factor & Dale 2014, Seitz et al. 2010). Potential factors are commonly grouped into "extrinsic" and "intrinsic" mechanisms (Lohr & Uhthoff 2007, Riley 2004, Seitz et al. 2010). Extrinsic mechanisms relate to factors acting from outside the tendon and include anatomical aspects such as acromial shape, subacromial impingement, internal impingement (i.e. impingement of

the rotator cuff against the superior glenoid rim), altered shoulder kinematics (e.g. glenohumeral instability), traumatic events or aspects of physical loading (e.g. overload). Intrinsic factors relate to vascular and morphological changes within the tendon, which may be caused e.g. by age (i.e. age-related degeneration), poor vascularity or genetics.

Regarding the different types of PTTs (see *section 2.2*), it has been proposed that bursal-sided PTT may primarily be associated with external mechanisms and articular-sided and intratendinous tears with intrinsic mechanisms (Seitz et al. 2010). Overall, though, the definitive aetiological relevance of these mechanisms and their interactions in the individual remain unclear.

The development of PTTs has commonly been viewed as part of a degenerative continuum (Factor & Dale 2014, Finnan & Crosby 2010, Matava et al. 2005). With his classification of "the progressive stages of [shoulder] impingement", Neer (1983) provided an early description of the progressive degenerative transition of pain-free and structurally intact tendons through to FTTs, which reflected his view of the pathogenesis at that time. His model encompassed three consecutive stages, each of which reflected specific changes within the tendon (Neer 1983 p. 72): "edema and haemorrhage" (stage 1), "fibrosis and tendinitis" (stage 2) and "bone spurs and tendon rupture" (stage 3). In the course of time, Neer's classification has been refined by him and others (e.g. Bigliani & Levine 1997, Cook & Purdam 2009, Jobe & Jobe 1983). In 2009, Cook & Purdam proposed "a new model of tendon pathology" (Cook & Purdam 2009 p. 410) based on a more contemporary understanding of the pathogenesis of rotator cuff degeneration. They also proposed three consecutive stages (*Figure 2.2*), which they described as follows (pp. 410-11): "reactive tendinopathy" (stage 1), defined as a "non-inflammatory proliferative response in the cell and matrix"; "tendon disrepair" (stage 2), defined as the "attempt at tendon healing, similar to reactive tendinopathy but with greater matrix breakdown"; and "degenerative tendinopathy" (stage 3), defined as "progression of both matrix and cell changes". The addition or removal of load is considered the primary stimulus for moving the tendon forward or backward along this continuum. Apart from a refinement of the pathophysiological processes, a notable change from Neer's model was the redesignation of tendinitis as "reactive tendinopathy", which was based on the view that rotator cuff tendinopathy is non-inflammatory. Histological studies had shown inflammatory cells to be absent in tendinopathy (Khan et al. 2002). This view has been challenged again by further research findings (Dean et al. 2016, Millar et al. 2016). The precise role of inflammation in tendon pathology, however, remains controversial and unclear (Rees 2016, Rees et al. 2014). Moreover, Cook and Purdam themselves, in a recently published review of

their model (Cook et al. 2016), acknowledge the complexities of the processes involved and state that “it is unlikely that any one model fully explains all aspects of the pathoaetiology of tendon pathology…” (p. 5).

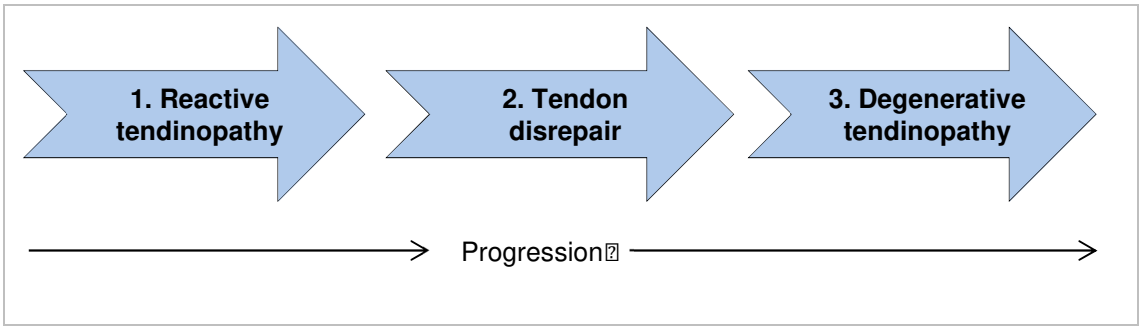


Figure 2.2: Outline of continuum of tendon pathology by Cook & Purdam (2009)

2.8 HEALING, NATURAL HISTORY AND TEAR PROGRESSION

2.8.1 HEALING OF PTTs

It is generally held that most PTTs do not heal spontaneously, though this is supported by little epidemiological and histological data (Fukuda 2003, Smith et al. 2010, Wolff et al. 2006). The putatively limited capacity for PTTs to heal has traditionally been attributed in part to *hypo*vascularity within the affected tendon, but this view has been challenged by more recent findings from in-vivo studies of *hyper*vascularity and hyperaemia at the site of pathology in symptomatic PTTs (Hegedus et al. 2010, Smith et al. 2010). In addition, in their follow-up studies of non-surgically treated PTTs, Maman et al. (2009) and Yamanaka and Matsumoto (1994) have found a decrease in tear size or even the complete disappearance of previously observed PTTs in small proportions (10%) of their populations. The issue of self-healing is thus contentious (Fukuda 2003).

2.8.2 NATURAL HISTORY AND PROGRESSION

Very little information is available on the natural history of painful PTTs in terms of their structural evolution and their clinical course over time. No data are available on the history of painful PTTs over a period with either no treatment or no formal treatment. Only two studies have followed populations of painful PTTs over a course of conservative treatment (Maman et al. 2009, Yamanaka & Matsumoto 1994). Both studies suggest an association between the incidence of progression of PTTs (same

as of FTTs) and increasing age (Maman et al. 2009, Yamanaka & Matsumoto 1994), but otherwise their findings are very discrepant. Maman et al. (2009) used MRI to evaluate 59 shoulders (in 54 participants, mean age 59 years) with painful rotator cuff tears (PTTs and FTTs) who were treated with physiotherapy, activity restriction and corticosteroid injections, and reported tear progression in 8% of the PTTs at a mean follow-up of 20 months. They presented no accompanying data on the course of pain and other symptoms, unfortunately. Yamanaka & Matsumoto (1994) evaluated 40 conservatively treated PTTs (36 participants, mean age 61 years; the content of treatment was unspecified) with arthrography, and noted that 80% had progressed at a mean follow-up (period between the initial and follow-up arthrography) of 412 days (approximately 14 months). They found that the incidence of progression was higher in initially larger PTTs and in those involving more than one tendon. On average (only the mean was presented), symptoms improved from 68 to 80 points on the Japanese Orthopaedic Association (JOA) score (0-100 points, 100 = best). In summary, the overall incidence of PTT progression is very uncertain, although the data suggest that it is not inevitable and that smaller PTTs and PTTs involving only one tendon may be less likely to progress, as may PTTs in younger patients. The course of symptoms over time either without or with conservative treatment is unknown.

2.8.3 POTENTIAL CONSEQUENCES OF TEAR PROGRESSION

Concerns about the progression of PTTs primarily relate to the development of FTTs, which can eventually cause significant permanent shoulder disability through extensive structural damage (Feeley et al. 2009, Kuzel et al. 2013, Laron et al. 2012). Mechanical detachment and retraction of the torn tear ends can induce irreversible fatty degeneration and atrophy of the affected muscles, developments which are associated with inferior outcomes of surgical repair (Kuzel et al. 2013, Laron et al. 2012). Moreover, longstanding massive rotator cuff tears can lead to glenohumeral arthritis and, as a rare but serious final stage, to cuff tear arthropathy, which is characterised by severe glenohumeral and acromioclavicular damage (Feeley et al. 2009, Neer et al. 1983).

2.9 DIAGNOSIS

The diagnosis of a painful PTT is usually based on clinical history, physical assessment and diagnostic imaging.

2.9.1 CLINICAL HISTORY AND PATIENT-REPORTED QUESTIONNAIRES

The purpose of the history in addition to evincing characteristics of impingement is to screen for other problems. The history-taking may involve patient-reported questionnaires, i.e. patient-reported outcome measures (PROMs), which are increasingly used to assess the impact of musculoskeletal symptoms on activities, participation and health-related quality of life (Vodicka et al. 2015). A large number of PROMs is available for use in shoulder pain populations (Roy & Esculier 2011, Wright & Baumgarten 2010). These can be divided into “upper extremity-specific”, “shoulder-specific” and “condition-specific” instruments (Roy & Esculier 2011 p. 341). Two rotator cuff-specific PROMs are available: the Rotator Cuff Quality of Life index (R-QoL) (Hollinshead et al. 2000) and the Western Ontario Rotator Cuff index (WORC) (Kirkley et al. 2003).

2.9.2 PHYSICAL ASSESSMENT

The physical assessment also includes a screening component for other causes of shoulder pain, such as cervical radiculopathy or frozen shoulder. The “impingement” element of the physical assessment involves tests to reproduce the pain of impingement (e.g. painful arc or Hawkins-Kennedy test) and tests to assess the painfulness and function of the rotator cuff muscles (isometric contractions and lag signs) (Hanchard et al. 2013, Hermans et al. 2013, Matava et al. 2005). Numerous physical tests have been proposed for diagnosing rotator cuff related disorders (Alqunaee et al. 2012, Hanchard et al. 2013, Hermans et al. 2013). There is, though, yet no agreement on the “best tests”, as most have been found to either lack accuracy or reliability or to be insufficiently researched (Hanchard et al. 2013, May et al. 2010).

2.9.3 DIAGNOSTIC IMAGING

Due to the largely nonspecific clinical presentation of painful PPTs, verification of the presence of a PTT ultimately requires diagnostic imaging. For this purpose, US and MRI are the most widely available and most commonly applied imaging techniques in clinical practice (Lenza et al. 2013, Smith et al. 2012). Magnetic resonance arthrography (MRA) is another option, but is invasive through intra-articular injection of a radiopaque contrast medium (Lenza et al. 2013). US and MRI are both noninvasive. A key advantage of US is that it allows dynamic visualisation of the rotator cuff during shoulder movements (Lenza et al. 2013). It is widely available, allowing for convenient, rapid assessment, and it is relatively inexpensive (Dinnes et al. 2003). It is also

generally covered by the (German) statutory health insurance. Generally accepted limitations of US are its long learning curve and related operator-dependence (Lenza et al. 2013). A key advantage of MRI is that it allows for high-resolution images in multiple planes (Dinnes et al. 2003). MRI has some contraindications, though, such as implanted ferromagnetic metallic devices; also, although MRI is widely viewed as a safe technique, the risks associated with exposure to the applied magnetic and radiofrequency fields have not yet been fully explored (Dill 2008). MRI further is more time-consuming and also more expensive (Lenza et al. 2013).

The evidence on the diagnostic accuracy of US, MRI and MRA for the detection of rotator cuff tears has been synthesised in various systematic reviews (de Jesus et al. 2009, Dinnes et al. 2003, Kelly & Fessell 2009, Lenza et al. 2013, Roy et al. 2015, Smith et al. 2011). *Table 2.1* provides summary data on the accuracy of US and MRI for both PTTs and FTTs from a recent comprehensive systematic review (Roy et al. 2015 Tables 2 & 3) which included overall 35 studies (2,774 shoulders) on US and 21 studies (1,575 shoulders) on MRI. The reference standard was surgery (either arthroscopic or open).

Table 2.1: Summary US and MRI accuracy statistics from Roy et al. (2015)

Type of tear	SN or SP	US	MRI
PTT	SN	68 (54 to 83)	67 (50 to 82)
	SP	94 (90 to 97)	94 (88 to 99)
FTT	SN	91 (86 to 94)	90 (85 to 95)
	SP	93 (91 to 96)	93 (89 to 97)

Values are % (95% CI); SN = sensitivity; SP = specificity

The data show that US and MRI have a similar diagnostic accuracy for the detection of PTTs and FTTs. Both have high specificity but, especially where PTTs are concerned, lower sensitivity. Thus, either imaging technique allows a PTT to be ruled in with a high degree of confidence, while neither allows a PTT to be confidently ruled out. Interobserver agreement regarding the detection of PTTs by US, when performed by experienced assessors, has been found to be high (Jeyam et al. 2008, Middleton et al. 2004).

2.10 TREATMENT

2.10.1 TREATMENT OPTIONS

Various conservative and surgical interventions are available for the treatment of PTTs (Huisstede et al. 2011, Seida et al. 2010). Both conservative and surgical treatment aim to alleviate pain and restore shoulder function and thereby aim to enable patients to achieve best-possible participation in their everyday life. They also aim to eliminate potential extrinsic causes for the PTT (e.g. by addressing altered scapular kinematics through physiotherapy or by arthroscopic subacromial decompression).

2.10.1.1 Conservative treatment

Conservative treatment most commonly encompasses advice (e.g. on temporary rest or activity modifications), physiotherapy, oral pain medication and/or subacromial injections (Finnan & Crosby 2010, Fukuda 2003, Wolff et al. 2006).

Physiotherapy

Physiotherapy offers a range of approaches and techniques (Finnan & Crosby 2010, Fukuda 2003, Wolff et al. 2006) that are essentially the same as those for subacromial impingement (Braun et al. 2013, Dong et al. 2015, Hanratty et al. 2012). Exercises are a mainstay of the treatment of patients with impingement-related shoulder pain (Bernhardsson et al. 2015, Johansson et al. 1999, Struyf et al. 2012); they may be supplemented by manual therapy and further modalities such as thermo- or electrotherapy. The precise biological responses to mechanical stimuli on tendon are yet insufficiently understood, but recent research suggests that controlled physical loading, i.e. in particular exercises, may play an important role in the stimulation of cellular proliferation and matrix remodelling, thereby promoting the regeneration of tendon and other musculoskeletal tissues (Huang et al. 2013, Khan & Scott 2009, Thompson et al. 2016). *Table 2.2* summarises common types and elements of the physiotherapy treatment for PTTs (i.e. impingement-related shoulder pain) (Bernhardsson et al. 2015, Johansson et al. 1999, Struyf et al. 2012), along with their main aims and assumed effects (Dölken 2005, van den Berg 2001). The table is not intended to present an exhaustive list and focuses on modalities that may be practiced by physiotherapists in Germany.

Table 2.2: Types and elements of physiotherapy for painful PTTs

Type	Elements	Main aims/assumed effects
Exercises	Coordinative exercises (e.g. complex movements)	To improve function through improvement of the coordinative abilities of the arm and shoulder
	Postural correction (e.g. spinal posture)	To improve pain and function by correcting/improving posture and to eliminate potential external factors fostering "impingement", e.g. excessive thoracic kyphosis
	Strengthening (e.g. concentric, eccentric, with/without equipment…)	To improve function in terms of the selective functioning of the affected or residual rotator cuff muscle(s) or of other shoulder or shoulder girdle muscles and to improve the overall functioning of the shoulder
	Scapular exercises (e.g. scapular positioning)	To improve function through normalisation or improvement of impaired scapular kinematics
	Stabilisation exercises (e.g. closed kinetic chain exercises)	To improve function through restoration or improvement of the stability of the glenohumeral joint
	Stretching exercises	To improve function through elimination or improvement of restrictions, e.g. shortened muscles, and to restore or improve ROM
Manual techniques	Manual mobilisations (articular)	To alleviate pain through application of mechanical stimuli to painful or restricted joints in the shoulder area (or at the cervical or thoracic spine), e.g. AC joint and to improve function by eliminating/improving movement restrictions
	Soft tissue mobilisations (e.g. deep transverse friction massage)	To alleviate pain and to decrease stiffness, i.e. increase flexibility/mobility of soft tissues at or around the shoulder
	Stretching (passive)	See stretching exercises
Electrotherapy	Therapeutic ultrasound	To reduce pain (and to promote structural regeneration and repair)
	Transcutaneous Electrical Nerve Stimulation (TENS)	To reduce pain
Thermotherapy	Cold (e.g. ice packs)	To reduce pain (and to modulate metabolic processes)
	Heat (e.g. hot packs)	To reduce pain and to enhance soft tissue flexibility and mobility (and to modulate metabolic processes)
Taping	Elastic taping, e.g. "kinesiotape"	To reduce pain and improve shoulder function by modulating muscle activity and soft tissue mobility

Injections and oral medication

The most commonly provided type of injections are subacromial corticosteroid injections, and the most common type of oral medication are nonsteroidal anti-inflammatory drugs (NSAIDs) (Finnan & Crosby 2010, Fukuda 2003, Wolff et al. 2006).

2.10.1.2 Surgical treatment

Surgical treatment options for PTTs include acromioplasty (subacromial decompression), debridement with or without acromioplasty or bursectomy, or tendon repair (Franceschi et al. 2012, Strauss et al. 2011). Surgery may be arthroscopic, mini-open or open, though arthroscopy appears to be the current standard (Franceschi et al. 2012). The rates of rotator cuff surgery, but in particular of arthroscopic surgery, have considerably increased internationally in recent years (Colvin et al. 2012, Judge et al. 2014, Svendsen et al. 2012, Yu et al. 2010).

2.10.2 EVIDENCE FOR THE EFFECTIVENESS OF DIFFERENT TREATMENT OPTIONS

The effectiveness of both conservative and surgical treatment approaches has been addressed by several recent systematic reviews on rotator cuff tears (Huisstede et al. 2011, Seida et al. 2010, Strauss et al. 2011) or rotator cuff tendinopathy (including PTTs) (Boudreault et al. 2014, Braun et al. 2013, Toliopoulos et al. 2014, van der Sande et al. 2013). Both have been shown to be effective in improving clinical outcomes. Due to a scarcity of follow-up studies on exclusive PTT populations, the evidence presented relates to mixed populations of impingement or rotator cuff tendinopathy, which, though, mostly incorporate PTTs.

2.10.2.1 Physiotherapy

The evidence for the effectiveness of exercises and manual therapy is detailed in *Chapter 4*.

The evidence for the effectiveness of supplementary physical modalities such as electrotherapies, thermotherapy or kinesio taping is overall limited and partly inconclusive (Desjardins-Charbonneau et al. 2015, Desmeules et al. 2015, Desmeules et al. 2016, Dong et al. 2015, Kromer et al. 2009, Yu et al. 2015). Most of these modalities appear to be either no more effective than placebo or of very limited benefit.

2.10.2.2 Corticosteroid injections and NSAIDs

The effectiveness of corticosteroid injections and NSAIDs for the treatment of rotator cuff disorders (including PTTs) appears unclear. Two systematic reviews (Coombes et al. 2010, van der Sande et al. 2013) found insufficient and conflicting evidence for the effectiveness of corticosteroid injections in improving outcomes such as pain and function compared to placebo or other interventions in both the short- and long-term. Based on limited evidence there is no difference in effectiveness between corticosteroid injections and physiotherapy (Foster 2015). Van der Sande et al. (2013) found a lack of evidence for the effectiveness of NSAIDs. According to another systematic review (Boudreault et al. 2014), oral NSAIDs may be effective in reducing pain in the short-term. The effectiveness of newly proposed injection therapies such as sodium hyaluronate injections remains to be established (Coombes et al. 2010).

2.10.2.3 Surgical treatment

Regarding the effectiveness of surgery for PTTs, a recent systematic review on the arthroscopic management of PTTs (Strauss et al. 2011) found widely varying rates (29% to 93%) of “excellent postoperative outcomes” (p. 573) across the 16 included studies. Limited evidence was found in favour of debridement with or without acromioplasty for smaller PTTs (involving < 50% of the tendon) and repair for larger PTTs (involving > 50% of the tendon). There was no evidence in favour of any specific repair technique. Postoperative complications were reported for 3% to 12% of cases and included post-surgical stiffness, persistent symptoms due to AC joint pathology, subcoracoid impingement and scapulothoracic bursitis. Further reported complications of arthroscopic rotator cuff repair include failure of repair, stiffness and infection (Randelli et al. 2012). Although shoulder arthroscopy is generally considered a safe procedure, it carries a small but real potential for life-threatening complications such as thromboembolism (Marecek & Saltzman 2010, Osti et al. 2012, Randelli et al. 2012).

2.10.2.4 Conservative versus surgical treatment

Direct comparisons of conservative versus surgical treatment as investigated in populations of impingement-related shoulder pain (Ketola et al. 2013, Saltychev et al. 2015) or FTTs (Kukkonen et al. 2015, Lambers Heerspink et al. 2015, Moosmayer et al. 2014) have mostly shown no clinically relevant differences in patient-centred outcomes such as pain or self-reported shoulder function. Compared with exercises, surgery has been found to be more expensive (Saltychev et al. 2015, Toliopoulos et al.

2014) and to involve more income transfers and sick leave (Toliopoulos et al. 2014). As yet, no such comparative study has investigated a specific population of patients with PTTs (Saltychev et al. 2015, Seida et al. 2010).

2.10.3 GUIDELINES AND INDICATIONS FOR TREATMENT

Current evidence-based guidelines on the management of rotator cuff tears (or impingement-related shoulder pain including PTTs) recommend conservative treatment (including various interventions) as the first-line treatment, with surgery to be considered where conservative treatment fails to yield satisfactory improvement (American Academy of Orthopaedic Surgeons (AAOS) 2010, Beaudreuil et al. 2010, Diercks et al. 2014). The optimal length of a trial of conservative treatment is unknown, with no specific recommendations given by the guidelines. Currently, no German national guideline on the management of rotator cuff disorders is available.

The recommendations of the guidelines reflect the considerable uncertainty about the specific indications for any treatment for PTTs, whether conservative or surgical. The often-mentioned “50% rule”, which asserts that PTTs involving 50% or more of tendon thickness should be treated surgically (Finnan & Crosby 2010, Fukuda 2003) lacks scientific support (Pedowitz et al. 2011). The prognosis of outcomes of treatment is addressed further below.

2.10.4 THE PATIENT PATHWAY IN THE GERMAN HEALTHCARE SYSTEM

In Germany, patients consult a medical practitioner in the first instance. This is usually a general practitioner (GP; “Hausarzt”), although patients are at liberty to self-refer to a specialist physician or surgeon. Medical practitioners may prescribe a course of physiotherapy. Self-referral to a physiotherapist is generally not allowed. Thus, a prescription is prerequisite for physiotherapy. For physiotherapy under the statutory health insurance system („Gesetzliche Krankenversicherung“, GKV), which covers 87% of the German population (GKV-Spitzenverband 2016), the allowable standard modalities and amount of physiotherapy for any complaint are regulated by the “Heilmittelrichtlinie” (“Healthcare directive”), which includes the “Heilmittelkatalog” (“Healthcare catalogue”) (GKV-Heilmittelrichtlinie 2011). PTTs generally fall under disease category “EX2” (“Injuries, operations and diseases of the limbs and pelvis”). The standard maximum number of treatment sessions is 18 (three sets of six sessions), the recommended frequency □ 2 sessions per week. Treatment may include “general physiotherapy” (“Allgemeine Krankengymnastik”), manual therapy (“Manuelle

□

Therapie") or therapeutic training ("Krankengymnastik am Gerät"), and a limited amount of supplementary modalities, e.g. electrotherapy or massage. Exceeding this limit may require the health insurer to approve a formal request. Physiotherapy under the private health insurance system is generally less regulated, but provision varies depending on the insurer and tariff. Injections and acupuncture are not within the scope of practice of physiotherapists in Germany.

Currently, no data are publicly available on the consumption of physiotherapy treatment for patients with rotator cuff tears in Germany. Meanwhile, the uptake of surgery appears to be increasing (see *section 2.6*).

2.11 PROGNOSIS AND PROGNOSIS RESEARCH

2.11.1 INTRODUCTION

It would benefit both patients and health care providers if likely responders and, by corollary, non-responders to conservative interventions, could be identified at the commencement of their care pathway. This would save time, effort and suffering, limit exposure to the risks of surgery and promote the optimal distribution of available resources.

In general, the importance of predicting which patients will respond to particular interventions is increasingly recognised and has stimulated a growing interest in prognosis and prognosis research (Croft et al. 2015, Moons et al. 2009a, Stanton et al. 2010). There has been a corresponding development in related prognosis research methodology (Cochrane Prognosis Methods Group 2016, Moons et al. 2009a, PROGRESS 2016). Prognosis in the context of clinical medicine is defined as "the risk of future health outcomes in people with a given disease or health condition" (PROGRESS_Research 2016), prognosis research consequentially as "the investigation of the relations between future outcomes ("endpoints") among people with a given baseline health state ("startpoint") in order to improve health" (PROGRESS_Research 2016). Estimates of prognosis are context-dependent, with relevant contextual factors being existing diagnostic and treatment practices, time and place. In the last few years, several initiatives have been established which focus on the contemporary advancement of prognosis research methodology; these include in particular the PROGnosis RESearch Strategy Partnership (PROGRESS 2016), a UK-based interdisciplinary collaboration of international researchers funded by the UK

Medical Research Council (MRC), and the Cochrane Prognosis Methods group (Cochrane Prognosis Methods Group 2016). It is widely acknowledged that prognosis research to date largely falls behind the high methodological standards of other areas of research (Altman 2009, Hayden et al. 2009, Hemingway 2006, Hemingway et al. 2009, Riley et al. 2007). Many prognostic studies in various areas of medicine and health have been found to show serious methodological deficiencies which critically affect the validity of their findings (Altman 2009, Hayden et al. 2009).

The design and conduct of this PhD programme of research was informed by relevant methodological publications which were available at its inception (e.g. Altman 2009, Altman et al. 2009, Moons et al. 2009a&b, Royston et al. 2009) and more recent publications relating to the work of or being proposed by PROGRESS (PROGRESS_Publications 2016) and the Cochrane Prognosis Methods Group (Cochrane Prognosis Methods Group 2016). In this thesis I use, as far as possible, the terminology recommended by PROGRESS, such as “prognosis research”, “prognostic factor”, “prognostic model research”, “startpoint” and “endpoint” (Hemingway et al. 2013, PROGRESS 2016).

The basic elements of prognosis research, as outlined by PROGRESS (adapted from Hemingway et al. (2013)), are shown in *Figure 2.3*.

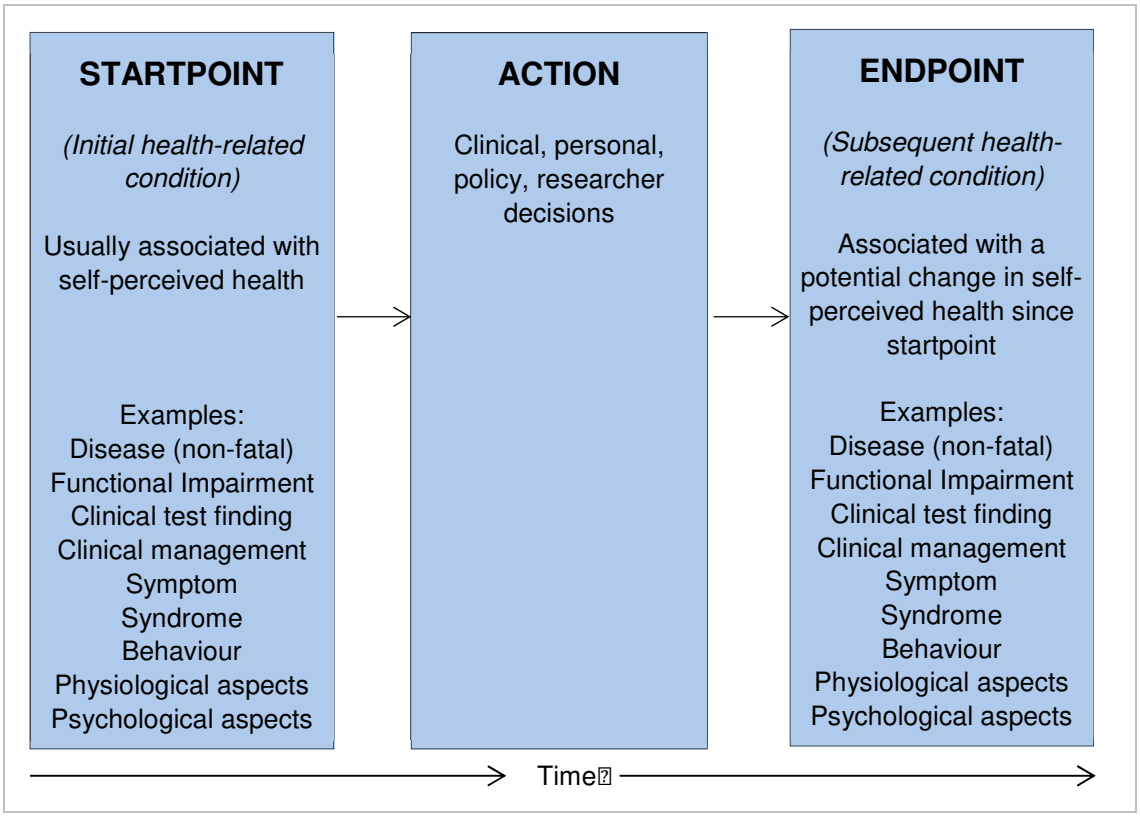


Figure 2.3: The basic elements of prognosis research *(note that estimates of prognosis are context-specific)*

PROGRESS have proposed a framework for prognosis research which encompasses four key interrelated research themes relating to the prognosis of clinical outcomes in people with a given health condition (Hemingway et al. 2013, Hingorani et al. 2013, Riley et al. 2013, Steyerberg et al. 2013). *Table 2.3* summarises the themes and provides an example for each based on a review on the prognosis of low back pain (Hayden et al. 2010).

Table 2.3: The PROGRESS framework for prognosis research

No	Theme	Description	Example
1	Fundamental prognosis research	The investigation of "the course of health related conditions in the context of the nature and quality of current care" (Hemingway et al. 2013), i.e. of the *average* prognosis of a subsequent clinical outcome in people with a given health related condition.	75-90% of patients presenting with an acute episode of low back pain for care will recover within a few weeks (Hayden et al. 2010).
2	Prognostic *factor* research	The investigation of prognostic factors, i.e. of „any measure that, among people with a given health condition, is associated with a subsequent clinical outcome." (PROGRESS_Research 2016)	Sciatica (referred leg pain) is associated with poor outcomes in acute or subacute low back pain (Hayden et al. 2010).
3	Prognostic *model* research	The investigation of prognostic models, i.e. of "multiple prognostic factors in combination to predict the risk of future clinical outcomes in individual patients" (PROGRESS_Research 2016)	The STarTBack Tool (Hill et al. 2008) aims to predict the risk (low, medium or high) of persistent symptoms and disability in patients with non-specific low back pain based on nine prognostic factors: bothersomeness, referred leg pain, comorbid pain, disability (two items), catastrophizing, fear, anxiety, and depression (Hayden et al. 2010).
4	Stratified medicine research	The investigation of targeting therapeutic decisions to subgroups of individuals based on the baseline prognosis of a clinical outcome (as established through prognostic factor or (ideally) prognostic model research).	Targeting treatment according to the risk (low, medium or high) of persistent symptoms and disability (based on the Start Back Tool, see above) improves outcomes in patients with low back pain (Hill et al. 2011).

2.11.2 PROGNOSTIC MODEL RESEARCH

The programme of research presented within this thesis focuses on prognostic *model* research (see *Chapter 1* for the primary research question). The key consideration for the choice of this type of prognosis research was the intention to predict the clinical outcome as accurately as possible by considering the real-life clinical complexities among individual patients. Prognostic models are best placed to do this because they account for interactions between multiple factors (Hemingway et al. 2013, Steyerberg et al. 2013). In the context of painful shoulder complaints, these typically include demographic or clinical factors, such as age, duration of symptoms or disability at baseline. The basic design of prognostic model research is outlined in *Figure 2.4*.

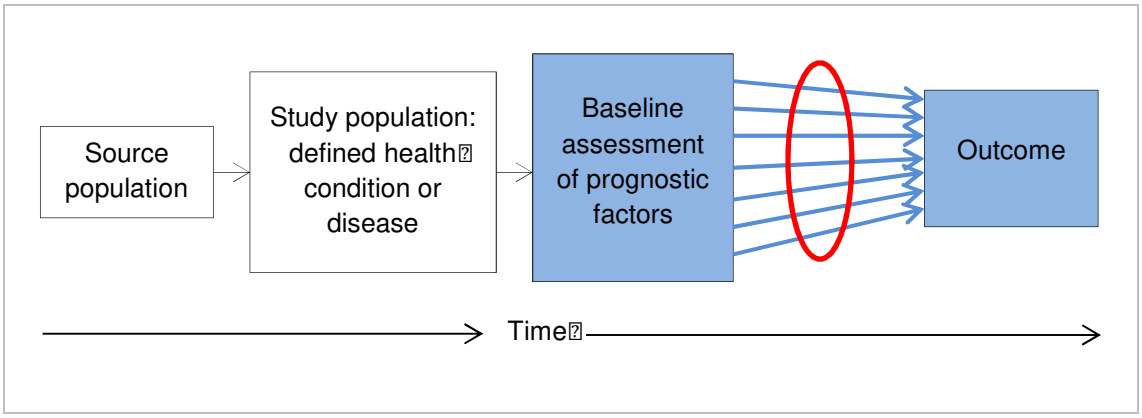


Figure 2.4: Basic design of prognostic model research
(adapted from 2014 PROGRESS prognosis research summer school, Keele University, UK, course materials)

Prognostic model research encompasses three consecutive key phases (*Figure 2.5*) (Harrell 2001, Steyerberg et al. 2013):

1) Model development: the initial determination of a prognostic model, which includes the internal validation of the model, i.e. the evaluation of its performance by use of data from the primary sample.
2) Model validation: the evaluation of the model's performance by use of data from independent samples, i.e. different clinical settings and populations. External validation is a crucial step before a model can be considered usable in clinical practice (Steyerberg et al. 2013).
3) Investigation of clinical impact: the evaluation of the model's effectiveness and cost-effectiveness in improving outcomes.

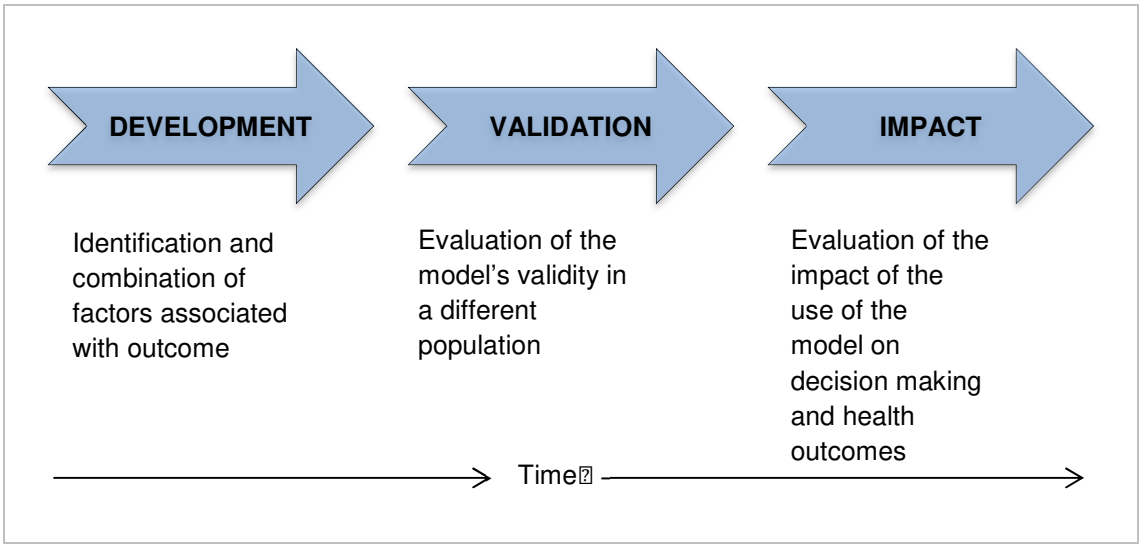


Figure 2.5: The phases of prognostic model research
(adapted from 2014 PROGRESS prognosis research summer school, Keele University, UK, course materials)

These three research phases show that the determination of a valid and usable prognostic model requires a comprehensive programme of research, from initial development through to the evaluation of clinical impact. The majority of published prognostic model studies are limited to model development, with few investigations of external validity and very few investigations of clinical impact (Altman 2009, Steyerberg & Harrell 2016, Steyerberg et al. 2013).

2.11.3 EVIDENCE ON THE PROGNOSIS OF OUTCOMES OF CONSERVATIVE TREATMENT OF ROTATOR CUFF DISORDERS

At the time when this programme of research was planned, only limited evidence was available on the prognosis of outcomes of conservative treatment with physiotherapy for rotator cuff disorders. In particular, no study was available that investigated a prognostic model for the outcome of a phase of treatment with physiotherapy in patients with painful PPTs, and no systematic review was available that specifically addressed prognostic models in rotator cuff disorders. The available evidence is addressed in *Chapter 3*.

2.12 RESEARCH GAPS AND NEEDS

In 2013, the US Agency for Healthcare Research and Quality (AHRQ) published a report of a comprehensive investigation of "future research needs" in the area of

conservative and surgical treatments for rotator cuff tears (Butler et al. 2013), which provided strong support of the relevance of the research presented in this thesis. This investigation encompassed an updated systematic review on interventions by Seida et al. (2010) and input from a group of stakeholders representing clinicians, researchers, professional organisations, research funders, payers and consumers. "Which treatment is best for which patient, and when?" was the „important overarching question" (p. 17) as identified by the stakeholders. "Understanding which patients do best with non-operative treatment" (p. 10) was rated as a top priority scientific research question. The main goal of consumers was to "return patients to full physical function" (p. 12). A top methods issue was "What is a minimally important difference in key outcomes?" (p. 9). The literature review also revealed that the majority of published research on the treatment for rotator cuff tears relates to surgical rather than non-surgical treatment. The initial review by Seida et al. (2010), which covered the literature from 1990 to 2009, had yielded a proportion of 82% of studies on surgical approaches or techniques (Butler et al. 2013).

2.13 SUMMARY AND CONTRIBUTIONS TO KNOWLEDGE

2.13.1 SUMMARY

In this chapter, I have provided the context and justification for the programme of research that is presented within the remainder of this thesis. The key aspects underpinning the relevance of the topic and field of research, i.e. of studying the prognosis of the outcome of a period of conservative treatment with physiotherapy in adults with painful PTTs, as presented in this and the precedent chapter (*Chapter 1*), are summarised in *Table 2.4*.

Table 2.4: Summary of background: underpinning the relevance of the topic and field of research

Aspect (related section)	**Key details**
Functional relevance of the rotator cuff (2.1)	The rotator cuff holds a primary role in the dynamic stabilisation of the glenohumeral joint.
Prevalence (1.1, 2.3, 2.7)	Most cases of shoulder pain involve the subacromial-subdeltoid bursa and the rotator cuff. The reported prevalence of PTTs in shoulder pain populations is up to 24%. The prevalence of PTTs, most of which are non-traumatic, is associated with increasing age.
Symptoms and burden of disease (2.4-2.6)	Painful rotator cuff tears can significantly affect shoulder function and health-related quality of life. People with PTTs may experience higher levels of pain than those with FTTs.
Healing, natural history and tear progression (2.8)	PTTs are unlikely to heal spontaneously and may progress over time. Progression may eventually lead to extensive structural damage to the shoulder.
Diagnosis (2.9)	US is highly specific for the detection of PTTs.
Treatment (2.10)	Patients with painful PTTs may profit from both conservative and surgical treatment. Conservative treatment (which usually includes physiotherapy) is recommended as the first-line treatment, but there is but there is insufficient knowledge on who responds (best) to what treatment. Thus, the precise indications for the different available treatment options are unclear.
Prognosis and prognostic research (2.11)	Being able to predict the clinical outcome of conservative treatment at the commencement of the patient's care pathway could save time, effort and suffering, limit exposure to the risks of surgery, and promote the optimal distribution of available resources.
	Prognostic models aim to predict future clinical outcomes in individuals by multiple factors in combination, thereby considering real-life clinical complexities among individual patients.
	There is a lack of both primary and secondary research on prognostic models for predicting the outcome of conservative treatment with painful PTTs. While PTTs are commonly included in study populations of impingement-related shoulder pain, there is a lack of research on exclusive PTT populations.
	Prognosis research methodology is developing and guidance is increasingly available.
Research gaps and needs (2.12)	A comprehensive investigation by (Butler et al 2013) yielded important issues and questions for research, including e.g.: • "Which treatment is best for which patient, and when?" ("important overarching question", p. 17) and • "Understanding which patients do best with non-operative treatment" (p. 10, top priority scientific research question), and • "What is a minimally important difference in key outcomes?" (top methods issue, p. 9)

2.13.2 CONTRIBUTIONS TO KNOWLEDGE

This programme of research was designed to provide substantial, original contributions to knowledge. The prognostic systematic review, the prognostic study, the MID analysis and the responder analysis (as outlined in *Chapter 1*) each represent original contributions to knowledge, as explained in the respective chapters.

2.14 REFERENCES

Alqunaee M., Galvin R., Fahey T. (2012). Diagnostic accuracy of clinical tests for subacromial impingement syndrome: a systematic review and meta-analysis. *Arch Phys Med Rehabil*, 93(2), 229-36. doi:10.1016/j.apmr.2011.08.035.

Altman D. G. (2009). Prognostic models: a methodological framework and review of models for breast cancer. *Cancer Invest*, 27(3), 235-43. doi:10.1080/07357900802572110.

Altman D. G., Vergouwe Y., Royston P., Moons K. G. M. (2009). Prognosis and prognostic research: validating a prognostic model. *BMJ*, May 28, 338, b605. doi:10.1136/bmj.b605.

American Academy of Orthopaedic Surgeons (AAOS). (2010). *American Academy of Orthopaedic Surgeons practice guideline: optimizing the management of rotator cuff problems. Guidelines and evidence report.* Available at http://www.aaos.org/research/guidelines/RCP_guideline.pdf [Last accessed 03 June 2016].

Bayam L., Ahmad M. A., Naqui S. Z., Chouhan A., Funk L. (2011). Pain mapping for common shoulder disorders. *Am J Orthop (Belle Mead NJ)*, 40(7), 353-8.

Beaudreuil J., Bardin T., Orcel P., Goutallier D. (2007). Natural history or outcome with conservative treatment of degenerative rotator cuff tears. *Joint Bone Spine*, 74(6), 527-9. doi:10.1016/j.jbspin.2007.07.009.

Beaudreuil J., Dhénain M., Coudane H., Mlika-Cabanne N. (2010). Clinical practice guidelines for the surgical management of rotator cuff tears in adults. *Orthop Traumatol Surg Res*, 96(2), 175-9. doi:10.1016/j.otsr.2010.02.002.

Bernhardsson S., Öberg B., Johansson K., Nilsen P., Larsson M. E. H. (2015). Clinical practice in line with evidence? A survey among primary care physiotherapists in western Sweden. *J Eval Clin Pract*, 21(6), 1169-77. doi:10.1111/jep.12380.

Bigliani L. U., Levine W. N. (1997). Subacromial impingement syndrome. *J. Bone Joint Surg Am*, 79(12), 1854-68.

Boudreault J., Desmeules F., Roy J.-S., Dionne C., Frémont P., Macdermid J. C. (2014). The efficacy of oral non-steroidal anti-inflammatory drugs for rotator cuff tendinopathy: a systematic review and meta-analysis. *J Rehabil Med*, 46(4), 294-306. doi:10.2340/16501977-1800.

Braun C., Bularczyk M., Heintsch J., Hanchard N. C. A. (2013). Manual therapy and exercises for shoulder impingement revisited. *Phys Ther Rev*, 18(4), 263-84. doi:10.1179/108331913X13709388114510.

Butler M., Forte M., Braman J., Swiontkowski M., Kane R. L. (2013). *Nonoperative and Operative Treatments for Rotator Cuff Tears: Future Research Needs. Identification of Future Research Needs From Comparative Effectiveness Review No. 22 (AHRQ Comparative Effectiveness Reviews)*. Future Research Needs Paper No 39. Agency for Healthcare Research and Quality (US). Available at: http://www.ncbi.nlm.nih.gov/books/NBK153178/ [Last accessed 03 June 2016].

Cadogan A., Laslett M., Hing W. a, McNair P. J., Coates M. H. (2011). A prospective study of shoulder pain in primary care: prevalence of imaged pathology and response to guided diagnostic blocks. *BMC Musculoskelet Disord*, 12(1), 119. doi:10.1186/1471-2474-12-119.

Cochrane Prognosis Methods Group. (2016). Cochrane Prognosis Methods Group. Available at: http://prognosismethods.cochrane.org/ [Last accessed 03 June 2016].

Colvin A. C., Egorova N., Harrison A. K., Moskowitz A., Flatow E. L. (2012). National trends in rotator cuff repair. *J Bone Joint Surg Am*, 94(3), 227-33. doi:10.2106/JBJS.J.00739.

Cook J. L., Purdam C. R. (2009). Is tendon pathology a continuum? A pathology model to explain the clinical presentation of load-induced tendinopathy. *Br J. Sports Med*, 43(6), 409-16. doi:10.1136/bjsm.2008.051193.

Cook J. L., Rio E., Purdam C. R., Docking S. I. (2016). Revisiting the continuum model of tendon pathology: what is its merit in clinical practice and research? *Br J Sports Med,* Apr 28 [Epub ahead of print]. doi:10.1136/bjsports-2015-095422.

Coombes B. K., Bisset L., Vicenzino B. (2010). Efficacy and safety of corticosteroid injections and other injections for management of tendinopathy: a systematic review of randomised controlled trials. *Lancet*, 6736(10), 1-17. doi:10.1016/S0140-6736(10)61160-9.

Croft P., Altman D. G., Deeks J. J., Dunn K. M., Hay A. D., Hemingway H., LeResche L., Peat G., Perel P., Petersen S. E., Riley R. D., Roberts I., Sharpe M., Stevens R. J., Van Der Windt D. a, Von Korff M., Timmis A. (2015). The science of clinical practice: disease diagnosis or patient prognosis? Evidence about “what is likely to happen” should shape clinical practice. *BMC Med*, 13(1), 20. doi:10.1186/s12916-014-0265-4.

Curry E. J., Matzkin E. E., Dong Y., Higgins L. D., Katz J. N., Jain N. B. (2015). Structural Characteristics Are Not Associated With Pain and Function in Rotator Cuff Tears: The ROW Cohort Study. *Orthop J Sport Med*, 3(5). doi:10.1177/2325967115584596.

de Jesus J. O., Parker L., Frangos A. J., Nazarian L. N. (2009). Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. *AJR Am J Roentgenol*, 192(6), 1701-7. doi:10.2214/AJR.08.1241.

Dean B. J. F., Gettings P., Dakin S. G., Carr A. J. (2016). Are inflammatory cells increased in painful human tendinopathy? A systematic review. *Br J Sports Med*, 50(4), 216-20. doi:10.1136/bjsports-2015-094754.

Desjardins-Charbonneau A., Roy J.-S., Dionne C. E., Desmeules F. (2015). The efficacy of taping for rotator cuff tendinopathy: a systematic review and meta-analysis. *Int J Sports Phys Ther*, 10(4), 420-33.

Desmeules F., Boudreault J., Roy J.-S., Dionne C., Frémont P., MacDermid J. C. (2015). The efficacy of therapeutic ultrasound for rotator cuff tendinopathy: A systematic review and meta-analysis. *Phys Ther Sport*, 16(3), 276-84. doi:10.1016/j.ptsp.2014.09.004.

Desmeules F., Boudreault J., Roy J. S., Dionne C. E., Frémont P., MacDermid J. C. (2016). Efficacy of transcutaneous electrical nerve stimulation for rotator cuff tendinopathy: a systematic review. *Physiotherapy*, 102(1), 41-9. doi:10.1016/j.physio.2015.06.004.

Diercks R., Bron C., Dorrestijn O., Meskers C., Naber R., de Ruiter T., Willems J., Winters J., van der Woude H. J. (2014). Guideline for diagnosis and treatment of subacromial pain syndrome. *Acta Orthop*, 85(3), 314-22. doi:10.3109/17453674.2014.920991.

Dill T. (2008). Contraindications to magnetic resonance imaging: non-invasive imaging. *Heart*, 94(7), 943-8. doi:10.1136/hrt.2007.125039.

Dinnes J., Loveman E., McIntyre L., Waugh N. (2003). The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review. *Health Technol Assess*, 7(29), iii, 1-166.

Dölken M. (2005). 'Überwiegend statisch bedingte Syndrome und Funktionskrankheiten' ['Primarly statically induced syndromes and functional disorders'], in Hüter-Becker A., Dölken M. (eds.) *Physiotherapie in der Orthopädie [Orthopaedic physical therapy]*. Stuttgart, Germany: Thieme, 213-8.

Dong W., Goost H., Lin X.-B., Burger C., Paul C., Wang Z.-L., Zhang T.-Y., Jiang Z.-C., Welle K., Kabir K. (2015). Treatments for shoulder impingement syndrome: a PRISMA systematic review and network meta-analysis. *Medicine (Baltimore)*, 94(10), e510. doi:10.1097/MD.0000000000000510.

Ejnisman B., Monteiro G. C., Andreoli C. V., de Castro Pochini A. (2010). Disorder of the long head of the biceps tendon. *Br J Sports Med*, 44(5), 347-54. doi:10.1136/bjsm.2009.064139.

Ellman H. (1990). Diagnosis and treatment of incomplete rotator cuff tears. *Clin Orthop Relat Res,* (254), 64-74.

Escamilla R. F., Yamashiro K., Paulos L., Andrews J. R. (2009). Shoulder muscle activity and function in common shoulder rehabilitation exercises. *Sports Med*, 39(8), 663-85. doi:10.2165/00007256-200939080-00004.

Factor D., Dale B. (2014). Current concepts of rotator cuff tendinopathy. *Int J Sports Phys Ther*, 9(2), 274-88.

Statistisches Bundesamt (2016). (Federal Statistical Office). *Diagnostic data of the hospitals starting from 2000 (key figures of full-time inpatients)*. Available at: http://www.gbe-bund.de (Search by topic: shoulder lesions Expenditures, Costs and Financing Diagnostic data of the hospitals…*;* document type: table) [Last accessed 03 June 2016].

Feeley B. T., Gallo R. A., Craig E. V. (2009). Cuff tear arthropathy: current trends in diagnosis and surgical management. *J Shoulder Elbow Surg*, 18(3), 484-94. doi:10.1016/j.jse.2008.11.003.

Finnan R. P., Crosby L. a. (2010). Partial-thickness rotator cuff tears. *J Shoulder Elbow Surg*, 19(4), 609-16. doi:10.1016/j.jse.2009.10.017.

Foster N. E. (2015). Similar clinical outcomes but more healthcare use in shoulder impingement patients following corticosteroid injection compared with physical therapy. *Evid Based Med*, 20(2), 67. doi:10.1136/ebmed-2015-110171.

Franceschi F., Papalia R., Palumbo A., Del Buono A., Maffulli N., Denaro V. (2012). Operative management of partial- and full-thickness rotator cuff tears. *Med Sport Sci*, 57, 100-13. doi:10.1159/000328888.

Fukuda H. (2000). Partial-thickness rotator cuff tears: a modern view on Codman's classic. *J Shoulder Elbow Surg*, 9(2), 163-8.

Fukuda H. (2003). The management of partial-thickness tears of the rotator cuff. *J. Bone Joint Surg Br*, 85(1), 3-11. doi:10.1302/0301-620X.85B1.13846.

Gartsman G. M., Milne J. C. (1995). Articular surface partial-thickness rotator cuff tears. *J Shoulder Elbow Surg*, 4(6), 409-15.

GKV-Heilmittelrichtlinie (2011). *Heilmittelrichtlinie und -katalog [Healthcare directive and catalogue]* (of the Gesetzliche Krankenversicherung [German statutory health insurance], GKV). Available at: https://www.gkv-spitzenverband.de/krankenversicherung/ambulante_leistungen/heilmittel/heilmittelrichtlinie/heilmittel-richtlinie.jsp [Last accessed 03 June 2016].

GKV-Spitzenverband (2016). [National Association of Statutory Health Insurance Funds (2013)]. Kennzahlen der gesetzlichen Krankenversicherung [Key figures of the statutory health insurance]. Available at: https://www.gkv-spitzenverband.de/presse/zahlen_und_grafiken/zahlen_und_grafiken.jsp [Last accessed 03 June 2016].

Gotoh M., Hamada K., Yamakawa H., Inoue A., Fukuda H. (1998). Increased substance P in subacromial bursa and shoulder pain in rotator cuff diseases. *J Orthop Res*, 16(5), 618-21. doi:10.1002/jor.1100160515.

Gschwend N., Ivosevi□-Radovanovi□ D., Patte D. (1988). Rotator cuff tear -relationship between clinical and anatomopathological findings. *Arch Orthop Trauma Surg*, 107(1), 7-15. doi:10.1016/0749-8063(89)90184-9.

Habermeyer P., Magosch P., Lichtenberg S. (2006). *Classifications and Scores of the Shoulder*. Berlin: Springer.

Hanchard N. C. A., Lenza M., Handoll H. H. G., Takwoingi Y. (2013). Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement. *Cochrane Database Syst Rev*, (4), CD007427. doi:10.1002/14651858.CD007427.pub2.

Hanratty C. E., McVeigh J. G., Kerr D. P., Basford J. R., Finch M. B., Pendleton A., Sim J. (2012). The effectiveness of physiotherapy exercises in subacromial impingement syndrome: a systematic review and meta-analysis. *Semin Arthritis Rheum*, 42(3), 297-316. doi:10.1016/j.semarthrit.2012.03.015.

Harrell F. E. (2001). *Regression modeling strategies*. New York: Springer.

Hayden J. A., Chou R., Hogg-Johnson S., Bombardier C. (2009). Systematic reviews of low back pain prognosis had variable methods and results: guidance for future prognosis reviews. *J Clin Epidemiol*, 62(8), 781-96.e1. doi:10.1016/j.jclinepi.2008.09.004.

□

Hayden J. A., Dunn K. M., van der Windt D. A., Shaw W. S. (2010). What is the prognosis of back pain? *Best Pract Res Clin Rheumatol*, 24(2), 167-79. doi:10.1016/j.berh.2009.12.005.

Hedtmann A. (2009). Weichteilerkrankungen der Schulter – Subakromialsyndrome. *Orthopädie und Unfallchirurgie up2date*, 4(02), 85-106. doi:10.1055/s-0029-1214456.

Heers G., Anders S., Werther M., Lerch K., Hedtmann a, Grifka J. (2005). [Efficacy of home exercises for symptomatic rotator cuff tears in correlation to the size of the defect]. *Sportverletz Sportschaden*, 19(1), 22-7. doi:10.1055/s-2004-813883.

Hegedus E. J., Cook C., Brennan M., Wyland D., Garrison J. C., Driesner D. (2010). Vascularity and tendon pathology in the rotator cuff: a review of literature and implications for rehabilitation and surgery. *Br J Sports Med*, 44(12), 838-47. doi:10.1136/bjsm.2008.053769.

Hemingway H. (2006). Prognosis research: why is Dr. Lydgate still waiting? *J Clin Epidemiol*, 59(12), 1229-38. doi:10.1016/j.jclinepi.2006.02.005.

Hemingway H., Riley R. D., Altman D. G. (2009). Ten steps towards improving prognosis research. *BMJ*, 339, b4184. doi:10.1136/bmj.b4184.

Hemingway H., Croft P., Perel P., Hayden J. a., Abrams K., Timmis A., Briggs A., Udumyan R., Moons K. G. M., Steyerberg E. W., Roberts I., Schroter S., Altman D. G., Riley R. D., PROGRESS Group. (2013). Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*, 346, e5595. doi:10.1136/bmj.e5595.

Hermans J., Luime J. J., Meuffels D. E., Reijman M., Simel D. L., Bierma-Zeinstra S. M. A. (2013). Does this patient with shoulder pain have rotator cuff disease?: The Rational Clinical Examination systematic review. *JAMA*, 310(8), 837–47. doi:10.1001/jama.2013.276187.

Hess S. A. (2000). Functional stability of the glenohumeral joint. *Man Ther*, 5(2), 63-71. doi:10.1054/math.2000.0241.

Hill J. C., Dunn K. M., Lewis M., Mullis R., Main C. J., Foster N. E., Hay E. M. (2008). A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum*, 59(5), 632-41. doi:10.1002/art.23563.

Hill J. C., Whitehurst D. G. T., Lewis M., Bryan S., Dunn K. M., Foster N. E., Konstantinou K., Main C. J., Mason E., Somerville S., Sowden G., Vohora K., Hay E. M. (2011). Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet*, 378(9802), 1560-71. doi:10.1016/S0140-6736(11)60937-9.

Hingorani A. D., Windt D. a. van der, Riley R. D., Abrams K., Moons K. G. M., Steyerberg E. W., Schroter S., Sauerbrei W., Altman D. G., Hemingway H., PROGRESS Group. (2013). Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ*, 346, e5793. doi:10.1136/bmj.e5793.

Hollinshead R. M., Mohtadi N. G., Vande Guchte R. a, Wadey V. M. (2000). Two 6-year follow-up studies of large and massive rotator cuff tears: comparison of outcome measures. *J Shoulder Elbow Surg*, 9(5), 373-81. doi:10.1067/mse.2000.108389.

Huang C., Holfeld J., Schaden W., Orgill D., Ogawa R. (2013). Mechanotherapy: Revisiting physical therapy and recruiting mechanobiology for a new era in medicine. *Trends Mol Med*, 19(9), 555-64. doi:10.1016/j.molmed.2013.05.005.

Huisstede B. M. A., Koes B. W., Gebremariam L., Keijsers E., Verhaar J. A.. (2011). Current evidence for effectiveness of interventions to treat rotator cuff tears. *Man Ther*, 16(3), 217-30. doi:10.1016/j.math.2010.10.012.

Ishii H., Brunet J. A., Welsh R. P., Uhthoff H. K. (1997). “Bursal reactions” in rotator cuff tearing, the impingement syndrome, and calcifying tendinitis. *J Shoulder Elbow Surg*, 6(2), 131-6. doi:10.1016/S1058-2746(97)90033-1.

Jeyam M., Funk L., Harris J. (2008). Are shoulder surgeons any good at diagnosing rotator cuff tears using ultrasound?: A comparative analysis of surgeon vs radiologist. *Int J Shoulder Surg*, 2(1), 4–6. doi:10.4103/0973-6042.39580.

Jobe F. W., Jobe C. M. (1983). Painful athletic injuries of the shoulder. *Clin Orthop Relat Res*, (173), 117-24. doi:10.1097/00003086-198303000-00015.

Johansson K., Adolfsson L., Foldevi M. (1999). Attitudes toward management of patients with subacromial pain in Swedish primary care. *Fam Pract*, *16*(3), 233-7.

Judge A., Murphy R. J., Maxwell R., Arden N. K., Carr A. J. (2014). Temporal trends and geographical variation in the use of subacromial decompression and rotator cuff repair of the shoulder in England. *Bone Joint J*, 96-B(1), 70-4. doi:10.1302/0301-620X.96B1.32556.

Kelly A. M., Fessell D. (2009). Ultrasound compared with magnetic resonance imaging for the diagnosis of rotator cuff tears: a critically appraised topic. *Semin Roentgenol*, 44(3), 196-200. doi:10.1053/j.ro.2009.03.013.

Ketola S., Lehtinen J., Rousi T., Nissinen M., Huhtala H., Konttinen Y. T., Arnala I. (2013). No evidence of long-term benefits of arthroscopicacromioplasty in the treatment of shoulder impingement syndrome: Five-year results of a randomised controlled trial. *Bone Joint Res*, 2(7), 132-9. doi:10.1302/2046-3758.27.2000163.

Khan K. M., Cook J. L., Maffulli N., Kannus P. (2000). Where is the pain coming from in tendinopathy? It may be biochemical, not only structural, in origin. *Br J Sports Med*, 34(2), 81-3.

Khan K. M., Cook J. L., Kannus P., Maffulli N., Bonar S. F. (2002). Time to abandon the “tendinitis” myth. *BMJ*, *324*(7338), 626-7.

Khan K. M., Scott a. (2009). Mechanotherapy: how physical therapists’ prescription of exercise promotes tissue repair. *Br J Sports Med*, 43(4), 247-52. doi:10.1136/bjsm.2008.054239.

Kirkley A., Alvarez C., Griffin S. (2003). The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. *Clin J Sport Med*, 13(2), 84-92.

Kromer T. O., Tautenhahn U. G., de Bie R. A., Staal J. B., Bastiaenen C. H. G. (2009). Effects of physiotherapy in patients with shoulder impingement syndrome: a systematic review of the literature. *J Rehabil Med*, 41(11), 870-80. doi:10.2340/16501977-0453.

Kukkonen J., Joukainen A., Lehtinen J., Mattila K. T., Tuominen E. K. J., Kauko T., Äärimaa V. (2015). Treatment of Nontraumatic Rotator Cuff Tears: A Randomized Controlled Trial with Two Years of Clinical and Imaging Follow-up. *J Bone Joint Surg Am*, 97(21), 1729-37. doi:10.2106/JBJS.N.01051.

Kuzel B. R., Grindel S., Papandrea R., Ziegler D. (2013). Fatty infiltration and rotator cuff atrophy. *J Am Acad Orthop Surg*, 21(10), 613-23. doi:10.5435/JAAOS-21-10-613.

Labriola J. E., Lee T. Q., Debski R. E., McMahon P. J. (2005). Stability and instability of the glenohumeral joint: the role of shoulder muscles. *J Shoulder Elbow Surg*, 14(1 Suppl S), 32S-8S. doi:10.1016/j.jse.2004.09.014.

Lambers Heerspink F. O., van Raay J. J. a M., Koorevaar R. C. T., van Eerden P. J. M., Westerbeek R. E., van 't Riet E., van den Akker-Scheek I., Diercks R. L. (2015). Comparing surgical repair with conservative treatment for degenerative rotator cuff tears: a randomized controlled trial. *J Shoulder Elbow Surg*, 24(8), 1274-81. doi:10.1016/j.jse.2015.05.040.

Laron D., Samagh S. P., Liu X., Kim H. T., Feeley B. T. (2012). Muscle degeneration in rotator cuff tears. *J Shoulder Elb Surg*, 21(2), 164-74. doi:10.1016/j.jse.2011.09.027.

Lashgari C., Redziniak D. (2012). The natural history of rotator cuff tears. *Curr Orthop Pract*, 23(1), 10-13.

Lenza M., Buchbinder R., Takwoingi Y., Johnston R. V, Hanchard N. C., Faloppa F. (2013). Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered. *Cochrane Database Syst Rev*, (9), CD009020. doi:10.1002/14651858.CD009020.pub2.

Leps C., Falz R., Sauer J., Bischoff C., Friedrich R., Schulze A., Thomas M., Busse M. (2012). Epidemiologische und ökonomische Bewertung von Muskel-Skelett-Erkrankungen für die ICD-10 Klasse M75 Schulterläsionen Epidemiological and economic valuation of musculoskeletal disorders for the ICD-10 class M75 Zusammenfassung. *Klin Sport*, 13(1), 1-6.

Lohr J. F., Uhthoff H. K. (2007). Epidemiology and pathophysiology of rotator cuff tears. *Orthopäde*, 36(9), 788-95.

Magarey M. E., Jones M. A. (2003). Dynamic evaluation and early management of altered motor control around the shoulder complex. *Man Ther*, 8(4), 195-206. doi:10.1016/S1356-689X(03)00094-8.

Mall N. A., Lee A. S., Chahal J., Sherman S. L., Romeo A. A., Verma N. N., Cole B. J. (2013). An evidenced-based examination of the epidemiology and outcomes of traumatic rotator cuff tears. *Arthroscopy*, 29(2), 366-76. doi:10.1016/j.arthro.2012.06.024.

Maman E., Harris C., White L., Tomlinson G., Shashank M., Boynton E. (2009). Outcome of nonoperative treatment of symptomatic rotator cuff tears monitored by magnetic resonance imaging. *J Bone Joint Surg Am*, 91(8), 1898–906. doi:10.2106/JBJS.G.01335.

Marecek G. S., Saltzman M. D. (2010). Complications in shoulder arthroscopy. *Orthopedics*, 33(7), 492-7. doi:10.3928/01477447-20100526-15.

Matava M. J., Purcell D. B., Rudzki J. R. (2005). Partial-thickness rotator cuff tears. *Am J Sports Med*, 33(9), 1405-17. doi:10.1177/0363546505280213.

May S., Chance-Larsen K., Littlewood C., Lomas D., Saad M. (2010). Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. *Physiotherapy*, 96(3), 179-90. doi:10.1016/j.physio.2009.12.002.

Middleton W. D., Teefey S. A., Yamaguchi K. (2004). Sonography of the rotator cuff: analysis of interobserver variability. *AJR Am J Roentgenol*, 183(5), 1465–8.

Millar N. L., Dean B. J., Dakin S. G. (2016). Inflammation and the continuum model: time to acknowledge the molecular era of tendinopathy. *Br J Sports Med,* Jun 3 [Epub ahead of print]. doi:10.1136/bjsports-2016-096419.

Minns Lowe C. J., Moser J., Barker K. (2014). Living with a symptomatic rotator cuff tear “bad days, bad nights”: a qualitative study. *BMC Musculoskelet Disord*, 15, 228. doi:10.1186/1471-2474-15-228.

Moons K. G. M., Royston P., Vergouwe Y., Grobbee D. E., Altman D. G. (2009a). Prognosis and prognostic research: what, why, and how? *BMJ*, *338*, b375. doi:10.1136/bmj.b375.

Moons K. G. M., Altman D. G., Vergouwe Y., Royston P. (2009b). Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*, *338*, b606–b606. doi:10.1136/bmj.b606.

Moosmayer S., Lund G., Seljom U. S., Haldorsen B., Svege I. C., Hennig T., Pripp A. H., Smith H.-J. (2014). Tendon repair compared with physiotherapy in the treatment of rotator cuff tears: a randomized controlled study in 103 cases with a five-year follow-up. *J Bone Jt Surg*, 96(18), 1504-14. doi:10.2106/JBJS.M.01393.

Neer C. S. (1983). Impingement lesions. *Clin Orthop Relat Res*, (173), 70-7.

Neer C. S., Craig E. V, Fukuda H. (1983). Cuff-tear arthropathy *J Bone Joint Surg Am*, 65(9), 1232-44.

Oh L. S., Wolf B. R., Hall M. P., Levy B. A., Marx R. G. (2007). Indications for rotator cuff repair: a systematic review. *Clin Orthop Relat Res*, 455, 52-63. doi:10.1097/BLO.0b013e31802fc175.

Osti L., Papalia R., Del Buono A., Denaro V., Maffulli N. (2012). Understanding and preventing complications in repairing rotator cuff tears. *Med Sport Sci*, 57, 178-83. doi:10.1159/000329812.

Pedowitz R. A., Higashigawa K., Nguyen V. (2011). The “50% rule” in arthroscopic and orthopaedic surgery. *Arthroscopy*, *27*(11), 1584-7. doi:10.1016/j.arthro.2011.06.014.

Petersen S. A., Murphy T. P. (2011). The timing of rotator cuff repair for the restoration of function. *J. Shoulder Elbow Surg.*, *20*(1), 62-8. doi:10.1016/j.jse.2010.04.045.

Piitulainen K., Ylinen J., Kautiainen H., Häkkinen A. (2012). The relationship between functional disability and health-related quality of life in patients with a rotator cuff tear. *Disabil. Rehabil.*, 34(24), 2071-5. doi:10.3109/09638288.2012.670363.

PROGRESS (2016). MRC PROGnosis RESearch Strategy Partnership. Available at: http://www.progress-partnership.org/ [Last accessed 03 June 2016].

PROGRESS_Publications (2016). *MRC PROGnosis RESearch Strategy Partnershipitle - Publications*. Available at: http://progress-partnership.org/publications/ [Last accessed 03 June 2016].

PROGRESS_Research (2016). *MRC PROGnosis RESearch Strategy Partnershipitle - Research*. *Available at:* http://progress-partnership.org/research/ [Last accessed 03 June 2016].

Randelli P., Spennacchio P., Ragone V., Arrigoni P., Casella A., Cabitza P. (2012). Complications associated with arthroscopic rotator cuff repair: a literature review. *Musculoskelet Surg*, 96(1), 9-16. doi:10.1007/s12306-011-0175-y.

Rees J. D., Stride M., Scott A. (2014). Tendons - time to revisit inflammation. *Br J Sports Med*, 48(21), 1553-7. doi:10.1136/bjsports-2012-091957.

Rees J. D. (2016). The role of inflammatory cells in tendinopathy: is the picture getting any clearer? *Br J Sports Med*, 50(4), 201-2. doi:10.1136/bjsports-2015-095174.

Reilly P., Macleod I., Macfarlane R., Windley J., Emery R. J. H. (2006). Dead men and radiologists don’t lie: a review of cadaveric and radiological studies of rotator cuff tear prevalence. *Ann R Coll Surg Engl*, 88(2), 116-21. doi:10.1308/003588406X94968.

Riley G. (2004). The pathogenesis of tendinopathy. A molecular perspective. *Rheumatology (Oxford)*, 43(2), 131-42. doi:10.1093/rheumatology/keg448.

Riley R. D., Ridley G., Williams K., Altman D. G., Hayden J., de Vet H. C. W. (2007). Prognosis research: toward evidence-based results and a Cochrane methods group. *J Clin Epidemiol*, 60(8), 863-5; author reply 865–6. doi:10.1016/j.jclinepi.2007.02.004.

Riley R. D., Hayden J. A., Steyerberg E. W., Moons K. G. M., Abrams K., Kyzas P. A., Malats N., Briggs A., Schroter S., Altman D. G., Hemingway H., PROGRESS Group. (2013). Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*, 10(2), e1001380. doi:10.1371/journal.pmed.1001380.

Roy J.-S., Esculier J.-F. (2011). Psychometric evidence for clinical outcome measures assessing shoulder disorders. *Phys Ther Rev*, 16(5), 331-346. doi:10.1179/1743288X11Y.0000000043.

□

Roy J.-S., Braën C., Leblond J., Desmeules F., Dionne C. E., MacDermid J. C., Bureau N. J., Frémont P. (2015). Diagnostic accuracy of ultrasonography, MRI and MR arthrography in the characterisation of rotator cuff disorders: a systematic review and meta-analysis. *Br J Sports Med*, *49*(20), 1316-28. doi:10.1136/bjsports-2014-094148.

Royston P., Moons K. G. M., Altman D. G., Vergouwe Y. (2009). Prognosis and prognostic research: Developing a prognostic model. *BMJ*, 338, b604–b604. doi:10.1136/bmj.b604.

Ryliskis S., Brophy R. H., Kocius M., Marx R. G. (2009). Shoulder activity level in the preoperative assessment of patients with rotator cuff tears. *Knee Surg Sports Traumatol Arthrosc*, 17(12), 1522-28. doi:10.1007/s00167-009-0904-z.

Saltychev M., Äärimaa V., Virolainen P., Laimi K. (2015). Conservative treatment or surgery for shoulder impingement: systematic review and meta-analysis. *Disabil Rehabil*, 37(1), 1-8. doi:10.3109/09638288.2014.907364.

Seida J. C., LeBlanc C., Schouten J. R., Mousavi S. S., Hartling L., Vandermeer B., Tjosvold L., Sheps D. M. (2010). Systematic review: nonoperative and operative treatments for rotator cuff tears. *Ann Intern Med*, 153(4), 246-55. doi:10.7326/0003-4819-153-4-201008170-00263.

Seitz A. L., McClure P. W., Finucane S., Boardman N. D., Michener L. a. (2010). Mechanisms of rotator cuff tendinopathy: Intrinsic, extrinsic, or both? *Clin Biomech (Bristol, Avon)*, 26(1), 1-12. doi:10.1016/j.clinbiomech.2010.08.001.

Shin K. M. (2011). Partial-thickness rotator cuff tears. *Korean J Pain*, 24(2), 69-73. doi:10.3344/kjp.2011.24.2.69.

Smith C. D., Corner T., Morgan D., Drew S. (2010). Partial thickness rotator cuff tears: what do we know? *Shoulder Elbow*, 2(2), 77-82. doi:10.1111/j.1758-5740.2010.00054.x.

Smith T. O., Back T., Toms A. P., Hing C. B. (2011). Diagnostic accuracy of ultrasound for rotator cuff tears in adults: a systematic review and meta-analysis. *Clin Radiol*, 66(11), 1036-48. doi:10.1016/j.crad.2011.05.007.

Smith T. O., Daniell H., Geere J.-A., Toms A. P., Hing C. B. (2012). The diagnostic accuracy of MRI for the detection of partial- and full-thickness rotator cuff tears in adults. *Magn Reson Imaging*, 30(3), 336-46. doi:10.1016/j.mri.2011.12.008.

Snyder S. J., Pachelli A. F., Del Pizzo W., Friedman M. J., Ferkel R. D., Pattee G. (1991). Partial thickness rotator cuff tears: results of arthroscopic treatment. *Arthroscopy*, 7(1), 1-7. doi:10.1016/0749-8063(91)90070-E.

Stanton T. R., Hancock M. J., Maher C. G., Koes B. W. (2010). Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. *Phys Ther*, 90(6), 843-54. doi:10.2522/ptj.20090233.

Steyerberg E. W., Moons K. G. M., van der Windt D. A., Hayden J. A., Perel P., Schroter S., Riley R. D., Hemingway H., Altman D. G. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med*, 10(2), e1001381. doi:10.1371/journal.pmed.1001381.

□

Steyerberg E. W., Harrell F. E. (2016). Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*, 69, 245-7. doi:10.1016/j.jclinepi.2015.04.005.

Strauss E. J., Salata M. J., Kercher J., Barker J. U., McGill K., Bach B. R., Romeo A. A., Verma N. N. (2011). The arthroscopic management of partial-thickness rotator cuff tears: a systematic review of the literature. *Arthroscopy*, 27(4), 568-80. doi:10.1016/j.arthro.2010.09.019.

Struyf F., De Hertogh W., Gulinck J., Nijs J. (2012). Evidence-based treatment methods for the management of shoulder impingement syndrome among dutch-speaking physiotherapists: An online, web-based survey. *J Manipulative Physiol Ther*, 35(9), 720-6. doi:10.1016/j.jmpt.2012.10.009.

Svendsen S. W., Frost P., Jensen L. D. (2012). Time trends in surgery for non-traumatic shoulder disorders and postoperative risk of permanent work disability: a nationwide cohort study. *Scand J Rheumatol*, 41(1), 59-65. doi:10.3109/03009742.2011.595375.

Tashjian R. Z. (2012). Epidemiology, natural history, and indications for treatment of rotator cuff tears. *Clin Sports Med*, 31(4), 589-604. doi:10.1016/j.csm.2012.07.001.

Thompson W. R., Scott A., Loghmani M. T., Ward S. R., Warden S. J. (2016). Understanding Mechanobiology: Physical Therapists as a Force in Mechanotherapy and Musculoskeletal Regenerative Rehabilitation. *Phys Ther*, 96(4), 560-9. doi:10.2522/ptj.20150224.

Toliopoulos P., Desmeules F., Boudreault J., Roy J. S., Frémont P., MacDermid J. C., Dionne C. E. (2014). Efficacy of surgery for rotator cuff tendinopathy: a systematic review. *Clin Rheumatol*, 33(10), 1373-83. doi:10.1007/s10067-014-2563-9.

van den Berg F. (2001). ‘Bewegungssystem’ [‘Movement system’]. In van den Berg F. *Angewandte Physiologie. Band 3. Therapie, Training, Tests [Applied Physiology. Volume 3. Therapy, Training, Tests]*. Stuttgart: Thieme, 1-254.

van Der Sande R., Rinkel W. D., Gebremariam L., Hay E. M., Koes B. W., Huisstede B. M. (2013). Subacromial impingement syndrome: Effectiveness of pharmaceutical interventions-nonsteroidal anti-inflammatory drugs, corticosteroid, or other injections: A systematic review. *Arch Phys Med Rehabil*, 94(5), 961-76.

Vodicka E., Kim K., Devine E. B., Gnanasakthy A., Scoggins J. F., Patrick D. L. (2015). Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007-2013). *Contemp Clin Trials*, 43, 1-9. doi:10.1016/j.cct.2015.04.004.

Wolff A. B., Sethi P., Sutton K. M., Covey A. S., Magit D. P., Medvecky M. (2006). Partial-thickness rotator cuff tears. *J Am Acad Orthop Surg*, 14(13), 715-25.

Wright R. W., Baumgarten K. M. (2010). Shoulder outcomes measures. *J Am Acad Orthop Surg*, 18(7), 436-44.

Yamaguchi K., Ditsios K., Middleton W. D., Hildebolt C. F., Galatz L. M. (2006). The demographic and morphological features of rotator cuff disease: a comparison of asymptomatic and symptomatic shoulders. *J Bone Joint Surg Am*, 88A(8), 1699-1704.

Yamanaka K., Matsumoto T. (1994). The joint side tear of the rotator cuff. A followup study by arthrography. *Clin Orthop Relat Res*, (304), 68-73.

Yu E., Cil A., Harmsen W. S., Schleck C., Sperling J. W., Cofield R. H. (2010). Arthroscopy and the dramatic increase in frequency of anterior acromioplasty from 1980 to 2005: An epidemiologic study. *Arthroscopy*, 26(9 SUPPL. 1), S142-7. doi:10.1016/j.arthro.2010.02.029.

Yu H., Cote P., Shearer H. M., Wong J. J., Sutton D. A., Randhawa K. A., Varatharajan S., Southerst D., Mior S. A., Ameis A., Stupar M., Nordin M., van der Velde G. M., Carroll L., Jacobs C. L., Taylor-Vaisey A. L., Abdulla S., Shergill Y. (2015). Effectiveness of Passive Physical Modalities for Shoulder Pain: Systematic Review by the Ontario Protocol for Traffic Injury Management Collaboration. *Phys Ther*, 95(3), 306-18. doi:10.2522/ptj.20140361.

PART TWO:

RESEARCH

Orientation Table Chapter 3

Part	Ch.	Title	Aims
ONE	1	General introduction, aims, content and structure of the thesis	1. To provide a general introduction to the topic 2. To summarise the aims, content and structure of the thesis
	2	Background	To provide the relevant topical and conceptual background to the programme of research
TWO	**3**	**Prognostic models in adults undergoing physiotherapy for rotator cuff disorders - a systematic review**	**To establish the state of evidence on prognostic models in adults undergoing physiotherapy for painful rotator cuff disorders**
	4	Developing and validating the physiotherapy protocol for the prognostic study	1. To establish the state of evidence on the effectiveness of physiotherapy interventions for adults with painful atraumatic PTTs 2. To develop and validate the physiotherapy treatment protocol
	5	Selecting and defining the candidate prognostic factors for the prognostic study	1. To identify and select the candidate factors for the prognostic model study 2. To define the specific measures for the selected factors
	6	Predicting the outcome of physiotherapy in adults with painful partial-thickness rotator cuff tears (PTTs) – a prognostic model study	To develop a prognostic model for the outcome of a period of conservative treatment with physiotherapy in adult patients with painful atraumatic PTTs
	7	Drawing meaningful magnitude-based inferences from the prognostic study – Minimal Important Difference (MID) and responder analysis	1. To establish an estimate of the MID of the Western Ontario Rotator Cuff Index (WORC), the primary outcome of the prognostic model study 2. To apply the estimated MID to an exploratory responder analysis
THREE	8	Overall summary and conclusions	1. To summarise the research 2. To provide overall conclusions and consider implications
FOUR		Appendices	Appendices to Chapters 3-7

CHAPTER 3
Prognostic models in adults undergoing physiotherapy for rotator cuff disorders – a systematic review

3.1 INTRODUCTION

As set out in the background chapter (*Chapter 2*), the precise treatment indications for PTTs remain unclear. “Understanding which patients [with rotator cuff tears] do best with nonoperative treatment” (Butler et al. 2013 p. 10) has been rated a high priority research issue. Multivariable prognostic models aim to predict clinical outcomes in individual patients (PROGRESS_Research 2016), thus enabling the identification of likely responders and, by corollary, nonresponders to (conservative) treatment. The availability of clinically usable prognostic models for predicting outcomes in people with painful PTTs would benefit patients for whom a course of conservative treatment with physiotherapy (the usual first-line approach) was being considered, potentially avoiding unnecessary delays and suffering, and reducing uncertainty and anxiety. Equally, where surgery is contemplated as a first resort, a usable model might limit unnecessary exposure to the associated small but serious risks (Osti et al. 2012).

Comprehensive literature searches, which I conducted up to March 2012 during the planning stage of my PhD, revealed a lack of research on prognostic models for predicting outcomes in adults undergoing conservative treatment with physiotherapy for painful PTTs. Further, there was no systematic review to synthesise the available evidence on prognostic models for the outcome of conservative treatment in adults with painful rotator cuff disorders.

3.2 AIMS

The aim was to systematically review the available evidence on primary studies exploring prognostic models for predicting clinical outcomes in adults with painful rotator cuff disorders undergoing conservative treatment with physiotherapy. This also aimed to provide context for my own prognostic model study.

This systematic review was published online in *Physical Therapy* in December 2015 (Braun et al. 2015). The reader is directed to *Appendix 3.1* for the full article, which comprises the review report and supplementary materials (*eTables 1-5*). This chapter

summarises the review and expands the discussion, with a focus on aspects related to risk of bias and the use of PROBAST (the instrument with which I assessed risk of bias and applicability of the included studies, see further). Additionally, reference is made to unpublished materials related to the review such as the review protocol, which are enclosed in the appendices to this chapter.

3.3 Methods

3.3.1 PROTOCOL AND REGISTRATION

The review was based on an a priori protocol (*Appendix 3.2*) and was registered in PROSPERO, the International Prospective Register of Systematic Reviews (PROSPERO 2016) (registration nr. CRD42014008973). Upon its completion (2 April 2014), and prior to the conduct of the review, the protocol was lodged with the chair of the Research Governance and Ethics Committee of the School of Health and Social Care at Teesside University (Dr Alasdair Macsween). Differences between the protocol and the review were documented (see *Appendix 3.1 eTable 1*).

3.3.2 CRITERIA FOR CONSIDERING STUDIES FOR INCLUSION

The review included primary studies, reported in English, exploring prognostic models, at any stage of prognostic model research, in adults undergoing physiotherapy, with or without other conservative measures, for painful rotator cuff disorders (any type). Primary outcomes were pain, shoulder disability, assessed via a validated PROM, and adverse events. Inclusion was limited to prospective investigations of prognostic factors elicited at the baseline assessment. Studies had to evaluate a prognostic model, i.e. multiple factors in combination. This practical criterion was given precedence over whether studies were strictly classifiable as prognostic model or prognostic factor studies. No restriction was placed on the type of multivariable analysis. A full description of the inclusion criteria are reported in the published review and also documented in the eligibility form (see further).

3.3.3 DATA SOURCES AND SEARCHES

The comprehensive search, which is documented in the published review (*Appendix 3.1 eTable 2*; the presented Medline search strategie was modified for the other

databases), included searches of Medline, Embase, Cinahl, Cochrane Central, PEDro and the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) from inception up to 19 October 2015. I supplemented the search with hand searches of the reference lists of relevant studies. Additionally, I matched the compilation of eligible studies with the findings from my previous searches.

3.3.4 STUDY SELECTION

Study selection was done independently by two people, i.e. me (CB) and Dr Hanchard (NH) or Dr Handoll (HH), with arbitration where disagreement persisted through involvement of a third person (NH, HH or Prof Batterham, AMB. We used a filter to assist the screening of titles and abstracts (*Appendix 3.3*) and a purpose-developed pilot-tested eligibility form to assist the screening of full texts (*Appendix 3.4*).

3.3.5 DATA EXTRACTION

Data extraction was done independently by two people (CB and NH), using two purpose-developed and pilot-tested data extraction forms: one for developmental studies (*Appendix 3.5*) and one for validation studies (*Appendix 3.6*). Based on our initial assessment of the poor quality of the eligible studies, we extracted only one prognostic model per study. This was either the reportedly final model or the most complete model including the main effects for all prognostic factors. We did not impute any data. Author contact was limited to the clarification of issues related to study eligibility.

3.3.6 ASSESSMENT OF RISK AND BIAS AND APPLICABILITY

Risk of bias and applicability were assessed independently by two people (CB and NH), with any persisting disagreement being resolved through involvement of a third person (AMB). We piloted early versions of PROBAST, the Prediction Study Risk of Bias Assessment Tool (Wolff et al. 2015), for the assessment of risk of bias and applicability of the included studies, and then redid the assessment using the latest version available provided by its lead developer, Dr Robert Wolff (version 10/09/2014, see *Appendix 3.7*)[4]. The final tool, which is similar to the one we used, is pending publication (personal communication with Dr Wolff, 23/05/2016).

[4]The inclusion of this PROBAST version in the thesis appendices was approved by Dr Wolff (personal communication, 17/12/2015).

PROBAST is designed to assess risk of bias and applicability of primary studies evaluating (developing or validating) prognostic models. It is domain-based, the five domains being: participant selection, predictors (i.e. prognostic factors), outcome, sample size and participant flow, and analysis. Each domain comprises a set of "signalling questions" to facilitate judgements about risk of bias: low, high, or unclear. Additionally, the first three domains are assessed for concerns (low, high, or unclear) about the applicability of the study's design and characteristics to the review question. A summative judgement across all domains leads to an overall rating of low, high or unclear risk of bias. Lastly, the usability of the model is rated as either "yes" or "no". We based the PROBAST assessment on an a priori developed coding manual (*Appendix 3.8*). At the time when the review was conducted, no formal guidance was available from the developers, but Dr Wolff assisted with the clarification of questions. The assessment was done with reference to the selected model (see *section 3.3.5*).

3.3.7 DATA SYNTHESIS AND ANALYSIS

The characteristics and results of all included studies were tabulated and narratively synthesised. In the absence of sufficient good-quality, comparable and externally validated models, we did not undertake quantitative data synthesis.

3.4 RESULTS

We screened the titles and abstracts of 5,899 articles and 54 full texts. We included five studies (Hallgren et al. 2014, Hung et al. 2010, Kromer et al. 2014, Merolla et al. 2011, Taheriazam et al. 2005). We obtained unpublished full multivariable model data relating to the trial of Hallgren et al. (2014). My own study and seven other potentially relevant ongoing studies were also identified (see *Appendix 3.1 eTable 4*).

3.4.1 SUMMARY OF THE CHARACTERISTICS OF INCLUDED STUDIES

All included studies were cohort studies, but in two studies the cohort was derived from pooled data from a randomised controlled trial (RCT). Four studies (Hallgren et al. 2014, Hung et al. 2010, Kromer et al. 2014, Taheriazam et al. 2005) were classified as model development, and one was reported as a validation study (Merolla et al. 2011).

Four studies (Hallgren et al. 2014, Hung et al. 2010, Kromer et al. 2014, Taheriazam et al. 2005) investigated mixed populations with impingement-related shoulder pain, and one study (Merolla et al. 2011) investigated a rotator cuff tear population (without

differentiating between PTTs and FTTs). Initial sample sizes ranged from 33 (Hung et al. 2010) to 102 (Hallgren et al. 2014, Taheriazam et al. 2005). Although varying in duration, content and dosage, physiotherapy was provided to all study participants; steroid injections were provided to all participants of one study (Hallgren et al. 2014) and were optional in another study (Taheriazam et al. 2005).

The studies were heterogeneous in many of their characteristics, such as the number of outcome events (for binary outcomes) or individuals (for continuous outcomes) (23 to 89) and the number of initially considered prognostic factors (eight to presumably > 60). Prognostic factors mainly involved demographic and clinical characteristics; one study (Kromer et al. 2014) investigated psychosocial factors. Each study used different outcome measures; the commonality, though, being that all used PROMs. Follow-up ranged from six weeks (Hung et al. 2010) to 12 months (Hallgren et al. 2014, Kromer et al. 2014, Taheriazam et al. 2005).

The methods for selecting prognostic factors for inclusion in the multivariable analysis, where specified, varied across the studies; two studies (Hung et al. 2010, Kromer et al. 2014) explicitly reported using an automated statistical method (e.g. analysis of univariable correlations between the prognostic factors and the outcome). Likewise, the approaches to multivariable modelling varied. Three studies (Hung et al. 2010, Kromer et al. 2014, Taheriazam et al. 2005) used an automated statistical method (e.g. stepwise regression).

Further details of the included studies are tabulated in the published report (*Appendix 3.1 Table 1 and eTable 5*).

3.4.2 RISK OF BIAS AND APPLICABILITY

The summary of the PROBAST ratings for all included studies is shown in *Table 3.1*. A more detailed table including the ratings for the signalling questions as part of the risk of bias assessment can be viewed in *Appendix 3.9*. All studies were rated as at high risk of bias, which was mainly due to issues within the PROBAST domains 3 to 5 (outcome, sample size and participant flow and analysis). The ratings were affected by numerous issues, namely: inclusion of prognostic factors in the outcome definition (Kromer et al. 2014, Merolla et al. 2011, Taheriazam et al. 2005); unclear or lack of blinding of outcome determination to prognostic factor information (Hung et al. 2010, Merolla et al. 2011); an unreasonable number (> 5) of prognostic factors in relation to the number of outcome events or individuals reported in the selected model (Hallgren et al. 2014, Merolla et al. 2011); unclear handling of missing data (Hallgren et al. 2014,

Hung et al. 2010, Kromer et al. 2014, Merolla et al. 2011, Taheriazam et al. 2005); use of univariable analyses to select prognostic factors (Hung et al. 2010, Kromer et al. 2014); unclear (Hallgren et al. 2014) or unspecified (Merolla et al. 2011) modelling methods; and failure to consider overfitting of data, complexities in the data, evaluation of performance measures or non-linear relationships (Hallgren et al. 2014, Hung et al. 2010, Kromer et al. 2014, Merolla et al. 2011, Taheriazam et al. 2005). Notably, the only validation study (Merolla et al. 2011) was at high risk of bias in most domains.

Concerns about applicability were rated as low for two studies, unclear for one study, and high for two studies (see *Table 3.1*) and mainly related to the PROBAST domain 2 ("predictors", otherwise referred to as prognostic factors in this thesis). Both risk of bias and applicability ratings were affected by inadequate reporting. We rated all models as not usable in clinical practice.

Table 3.1: PROBAST (risk of bias and applicability) ratings

Domain/ Study ID	Risk of Bias					Applicability concerns			Overall Judgements		
	1. Participant Selection	2. Predictors	3. Outcome	4. Sample Size & Flow	5. Analysis	1. Participant Selection	2. Predictors	3. Outcome	Risk of Bias	Applicability	Usability
Björnsson Hallgren 2014	□	□	□	□	?	□	□	□	□	□	□
Hung 2010	□	?	□	?	□	?	□	?	□	□	□
Kromer 2014	□	□	□	?	□	□	□	□	□	□	□
Merolla 2011	?	?	□	□	?	?	□	?	□	□	□
Taheriazam 2005	□	□	□	?	?	?	?	□	□	?	□

ID = first author, yr; □ = *low risk/concerns;* ? = *unclear risk/concerns;* □ = *high risk/concerns*

3.4.3 RESULTS OF INCLUDED STUDIES

The presented models were heterogeneous in various aspects, including the number and composition of prognostic factors, and in terms of the presented statistics. A summary table presenting the characteristics and results of the studies is available in

□

the supplementary materials to the published report (*Appendix 3.1 eTable 5*). None of the four developmental studies reported any form of model validation, and none of these studies were followed by an external validation. Finally, none of the models evaluated in the included studies were assessed for clinical impact.

3.5 DISCUSSION

This systematic review included five studies (387 participants) which aimed to either develop or validate prognostic models for predicting outcomes in adults undergoing physiotherapy, with or without other conservative measures, for painful rotator cuff disorders. The studies were heterogeneous in many characteristics. The heterogeneity ruled out meaningful quantitative synthesis and imposed major limitations on the narrative synthesis. All included studies were rated as at high risk of bias, and most raised unclear or high concerns about applicability. The assessment of the studies was affected by reporting deficiencies. We considered none of the five models as usable in practice.

3.5.1 APPLICABILITY

The study populations were broadly relevant to the review question. Four studies investigated populations with impingement-related shoulder pain which implicitly included rotator cuff tears of any completeness, except for one study (Hung et al. 2010) which excluded FTTs. One study (Merolla et al. 2011) exclusively studied rotator cuff tears, although it is unclear whether PTTs were included. Applicability was compromised by unclear eligibility criteria in some studies, pertaining, for example, to frozen shoulder (Hung et al. 2010) or rotator cuff tears (Hallgren et al. 2014, Kromer et al. 2014, Taheriazam et al. 2005). In two studies, the patient populations were selected by virtue of their agreement to participate in an RCT (Hallgren et al. 2014, Kromer et al. 2014), which may have reduced external validity. No study investigated an exclusive population of people with PTTs.

The physiotherapy treatment was generally consistent with standard clinical practice; it was insufficiently reported to allow for a judgement in one study (Taheriazam et al. 2005).

The selection of prognostic factors was diverse and generally unjustified. Applicability was compromised by various issues. In one study (Hung et al. 2010), specialised equipment (a special motion analysis system) was used which would not be available

in most clinical settings. The reproducibility of some of the models is likely to be compromised by the questionable measurement properties of some prognostic factor measurements (e.g. posterior shoulder tightness in Hung et al. (2010) and categorisation of continuous prognostic factors (Hallgren et al. 2014, Hung et al. 2010, Merolla et al. 2011, Taheriazam et al. 2005).

Of our pre-specified outcomes, the following were reported by the included studies: pain (Merolla et al. 2011), shoulder disability (Kromer et al. 2014, Merolla et al. 2011, Taheriazam et al. 2005), global perceived change (Hung et al. 2010) and need for surgery (Hallgren et al. 2014).

The lack of validation of any models in the four developmental studies (Hallgren et al. 2014, Hung et al. 2010, Kromer et al. 2014, Taheriazam et al. 2005), as well as of any investigation of clinical impact, presents a major obstacle to the usability of the models.

3.5.2 RISK OF BIAS

We evaluated risk of bias in five domains: participant selection, predictors (i.e. prognostic factors), outcome, sample size and flow, and analysis. Various methodological issues affected our judgement of risk of bias (see results section). Of these, some issues, especially those relating to the number of outcome events or individuals in relation to the number of prognostic factors and the use of univariable analysis to select prognostic factors, have been shown to result in biased and unreliable models (Harrell et al. 1996).

Prognostic models have been shown to produce overoptimistic (i.e. exaggerated) predictions under a number of conditions. One occurs where samples have a small number of outcome events or individuals per studied prognostic factor (Harrell 2001 pp. 60-1, Moons et al. 2014, Steyerberg et al. 2013). Others include selection of factors for inclusion in the multivariable analysis based on the statistical significance of their univariable associations with the outcome (Bouwmeester et al. 2012, Harrell et al. 1996, Royston et al. 2009); and selection of factors within the multivariable analysis by automated procedures which rely entirely on statistical significance testing, such as stepwise regression (Flom & Cassell 2007, Harrell 2001 pp. 56–8, Miles & Shevlin 2001 pp. 38-9). Given that one or more of these aspects applied to all four developmental studies (Hallgren et al. 2014, Hung et al. 2010, Kromer et al. 2014, Taheriazam et al. 2005), the presented models are highly unlikely to produce valid and reliable predictions. The fifth study (Merolla et al. 2011), although reportedly a validation study, was seriously flawed in both concept and design.

□

Furthermore, the categorisation, but in particular the dichotomisation, of continuous prognostic factors has been found to produce various problems, such as a considerable loss of information, the consequential reduction of statistical power to detect a relationship between the prognostic factor and the outcome and the underestimation of the variation in the outcome between the groups (Altman 2006, Royston et al. 2006). Thus, the models of the four studies in which continuous prognostic factors were categorised (Hallgren et al. 2014, Hung et al. 2010, Merolla et al. 2011, Taheriazam et al. 2005) are likely to be biased.

Deficiencies such as unclear handling of missing data and the failure to consider overfitting of data, complexities in the data, evaluation of performance measures or non-linear relationships hampered the judgement of the quality of the data and of the models' performance.

One key issue is the inclusion of prognostic factors in the outcome definition. This is the problem of incorporation bias through mathematical coupling, and represents a conflict between risk of bias and applicability. The literature on incorporation bias primarily relates to diagnostic research, where it relates to the interaction between index and references tests (Reitsma et al. 2009). Mathematical coupling, which inherently occurs "when one variable directly or indirectly contains the whole or part of another" (Tu & Gilthorpe 2007 p. 444), may either erroneously purport a relationship between the prognostic factors and the outcome or overestimate an existing relationship. The conflict with applicability arises specifically because baseline and endpoint evaluation of a given outcome measure is standard clinical practice. This approach particularly relates to the increased use of patient-reported outcome measures (PROMs) in clinical practice and research (Vodicka et al. 2015). Moreover, in the present context, PROMs are among the very few prognostic factors that have a basis in evidence (Chester et al. 2013, Kuijpers et al. 2004). This conflict was encountered in two studies (Kromer et al. 2014, Taheriazam et al. 2005) which were both downgraded for risk of bias in the outcome domain as no adjustments were made in their study design or analysis to address incorporation bias.

3.5.3 POTENTIAL BIASES IN THE REVIEW PROCESS

This systematic review was based on an a priori protocol which was registered with PROSPERO. The protocol was lodged with the chair of the Research Governance and Ethics Committee of the School of Health and Social Care at Teesside University. Any deviations from the protocol were documented.

Although failure to identify relevant studies, especially unpublished articles or those in non-indexed journals, cannot be ruled out, the searches were comprehensive and included several supplementary sources. The yield of included studies from the initial search results was < 0.1%, which illustrates the known difficulties with the identification of prognosis research (Chatterley & Dennett 2012, Geersing et al. 2012) which include the lack of appropriate indexing functions in databases and of current validated search filters. I identified several search filters for prognosis research (e.g. Altman 2001, Geersing et al. 2012, Walker-Dilks et al. 2008, Wilczynski & Haynes 2004, 2005). However, I had concerns about the currentness of all but one (Geersing et al. 2012) which was purposely designed to identify prognostic model studies for systematic reviews. The use of this search filter appeared to be very helpful as it significantly decreased the number of results in Medline but still retrieved all five studies that were included in this review.

The identification of studies was further hampered by uninformative titles and abstracts and use of inconsistent terminology, as has been noted by others (Hemingway et al. 2013, Steyerberg et al. 2013). This lack of clarity similarly applied to the formulation of the developmental studies' objectives, which made it difficult to establish whether they intended to develop a prognostic model or whether they were interested in the identification of individual factors. In anticipation of this difficulty, any study was considered for inclusion in which two or more factors of interest were analysed in combination within the multivariable modelling.

Although inclusion of studies was limited to reports in English, there was no language restriction applied in the searches. Nonetheless, I did not identify any non-English but clearly relevant studies. The identified ongoing studies provide an indication of potentially relevant studies that may be published in the nearer future, although, due to insufficient provision of details in the majority, the eventual relevance of most of them to the question of this review is unclear.

Systematic reviewing of prognostic studies is an evolving field, and the methodology is work in progress. Nonetheless, I designed and conducted this review to the best possible standards, as far as possible: thus, in consideration of the contemporary guidance available from the PROGRESS partnership (PROGRESS 2016) and relevant recent methodological publications. This included evaluating the included studies using PROBAST, which is the first instrument that has specifically been designed to assess risk of bias and applicability of prognostic model studies. PROBAST was in development during the conduct of the systematic review, and my contact with the lead developer of the tool, Robert Wolff, resulted in my involvement in pilot-testing several

versions. This led to mutually useful discussions with Dr Wolff, who was generous in providing support and feedback on the use of PROBAST throughout. Moreover, Dr Wolff further shared my questions and feedback with the members of the PROBAST steering group, which, according to Dr. Wolff, helped to further improve the tool; examples are the inclusion of “justification for risk of bias (and applicability)” fields, improvements of the wording of the signalling questions and the identification of aspects that require specification or clarification in the background to the tool (personal communication with Dr Wolff, 13/01/2016).

Clearly, using an intermediate version of PROBAST, together with our own coding manual in the absence of a guidance document, means that there may be some variation in our interpretation and judgement of some of the items from that resulting from the use of the final version of the tool. This is unlikely to be a problem and anyway, since systematic reviewing of prognostic model studies is an evolving field and the methodology is work in progress, increasing knowledge is likely to affect the critical assessment of prognostic model studies.

3.5.4 AGREEMENT AND DISAGREEMENT WITH OTHER SYSTEMATIC REVIEWS

To my knowledge, this is the first systematic review of the evidence on primary prognostic model studies in adults with painful rotator cuff disorders undergoing conservative treatment with physiotherapy. I identified two other prognostic systematic reviews addressing shoulder pain (Chester et al. 2013, Kuijpers et al. 2004), both of which, though, aimed to synthesise evidence on individual prognostic factors rather than on prognostic models. Both reviews did not provide any subgroup analyses to allow for inferences about rotator cuff disorders.

3.5.5 IMPLICATIONS FOR PRACTICE AND RESEARCH

The systematic review demonstrated that there is no prognostic model ready to inform clinical practice on the prognosis of outcomes in adults undergoing physiotherapy, with or without other conservative measures, for painful rotator cuff disorders. It thus endorsed the need for my own primary study.

The complexity of prognostic modelling demands a high level of methodological expertise and clinical judgement, but particularly calls for the involvement, from the outset, of a statistician with expertise in the field. The composition of both primary and

secondary research teams should reflect this need. Researchers should be receptive to developing methodologies. Crucially, more attention should be paid to model validation, and ultimately, to the assessment of clinical impact.

3.6 REFERENCES

Altman D. G. (2001). Systematic reviews of evaluations of prognostic variables. *BMJ*, 323(7306), 224-8. doi:10.1136/bmj.323.7306.224.

Altman D. G. (2006). The cost of dichotomising continuous variables. *BMJ*, 332(7549), 1080. doi:10.1136/bmj.332.7549.1080.

Bouwmeester W., Zuithoff N. P. A., Mallett S., Geerlings M. I., Vergouwe Y., Steyerberg E. W., Altman D. G., Moons K. G. M. (2012). Reporting and methods in clinical prediction research: A systematic review. *PLoS Med*, 9(5), 1-12. doi:10.1371/journal.pmed.1001221.

Braun C., Hanchard N. C., Batterham A. M., Handoll H. H., Betthäuser A. (2015). Prognostic Models in Adults Undergoing Physical Therapy for Rotator Cuff Disorders: Systematic Review. *Phys Ther*, Dec 4 [Epub ahead of print]. doi:10.2522/ptj.20150475.

Butler M., Forte M., Braman J., Swiontkowski M., Kane R. L. (2013). *Nonoperative and Operative Treatments for Rotator Cuff Tears: Future Research Needs. Identification of Future Research Needs From Comparative Effectiveness Review No. 22 (AHRQ Comparative Effectiveness Reviews)*. Future Research Needs Paper No 39. Agency for Healthcare Research and Quality (US). Available at: http://www.ncbi.nlm.nih.gov/books/NBK153178/ [Last accessed 03 June 2016].

Chatterley T., Dennett L. (2012). Utilisation of search filters in systematic reviews of prognosis questions. *Health Info Libr J*, 29(4), 309-22. doi:10.1111/hir.12004.

Chester R., Shepstone L., Daniell H., Sweeting D., Lewis J., Jerosch-Herold C. (2013). Predicting response to physiotherapy treatment for musculoskeletal shoulder pain: a systematic review. *BMC Musculoskelet Disord*, 14(1), 203. doi:10.1186/1471-2474-14-203.

Flom, P.L., Cassell D. L. (2007). Statistics and Data Analysis: Why stepwise and similar selection methods are bad , and what you should use (statistics and data analysis). In *North East SAS Users Group (NESUG) 2007 Annual Conference*. Renaissance Harborplace Hotel, Baltimore, USA: SAS. Available at: http://www.lexjansen.com/pnwsug/2008/DavidCassell-StoppingStepwise.pdf [Last accessed 16 June 2016].

Geersing G.-J., Bouwmeester W., Zuithoff P., Spijker R., Leeflang M., Moons K. G. M., Moons K. (2012). Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS One*, 7(2), e32844. doi:10.1371/journal.pone.0032844.

Hallgren H. C. B., Holmgren T., Oberg B., Johansson K., Adolfsson L. E. (2014). A specific exercise strategy reduced the need for surgery in subacromial pain patients. *Br J Sports Med*, 48(19), 1431-6. doi:10.1136/bjsports-2013-093233.

Harrell F. E., Lee K. L., Mark D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*, 15(4), 361-87. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

Harrell F. E. (2001). *Regression modeling strategies*. New York: Springer.

Hemingway H., Croft P., Perel P., Hayden J. A., Abrams K., Timmis A., Briggs A., Udumyan R., Moons K. G. M., Steyerberg E. W., Roberts I., Schroter S., Altman D. G., Riley R. D., for the Progress Group. (2013). Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ*, 346, e5595. doi:10.1136/bmj.e5595.

Hung C.-J., Jan M.-H., Lin Y.-F., Wang T.-Q., Lin J.-J. (2010). Scapular kinematics and impairment features for classifying patients with subacromial impingement syndrome. *Man Ther*, 15(6), 547-51. doi:10.1016/j.math.2010.06.003.

Kromer T. O., Sieben J. M., de Bie R. A., Bastiaenen C. H. G. (2014). Influence of fear-avoidance beliefs on disability in patients with subacromial shoulder pain in primary care: a secondary analysis. *Phys Ther*, 94(12), 1775-84. doi:10.2522/ptj.20130587.

Kuijpers T., van Der Windt D. A. W. M., Van Der Heijden G. J. M. G., Bouter L. M. (2004). Systematic review of prognostic cohort studies on shoulder disorders. *Pain*, 109(3), 420-31. doi:10.1016/j.pain.2004.02.017.

Merolla G., Paladini P., Saporito M., Porcellini G. (2011). Conservative management of rotator cuff tears: literature review and proposal for a prognostic prediction score. *Muscles Ligaments Tendons J*, 1(1), 12-9.

Miles, Jeremy; Shevlin M. (2001). *Applying regression & correlation*. London, UK: Sage Publications.

Moons K. G., de Groot J. A., Bouwmeester W., Vergouwe Y., Mallett S., Altman D. G., Reitsma J. B., Collins G. S. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS Checklist. *PLoS Med*, 11(10), e1001744. doi:10.1371/journal.pmed.1001744.

Osti L., Papalia R., Del Buono A., Denaro V., Maffulli N. (2012). Understanding and preventing complications in repairing rotator cuff tears. *Med Sport Sci*, 57, 178-83. doi:10.1159/000329812.

PROGRESS (2016). MRC PROGnosis RESearch Strategy Partnership. Available at: http://www.progress-partnership.org/ [Last accessed 03 June 2016].

PROGRESS_Research (2016). MRC PROGnosis RESearch Strategy Partnershipitle - Research. Available at: http://progress-partnership.org/research/ [Last accessed 03 June 2016].

PROSPERO (2016). International Prospective Register of Systematic Reviews. Available at: http://www.crd.york.ac.uk/PROSPERO/ [Last accessed 03 June 2016].

Reitsma J. B., Rutjes A. W. S., Whiting P., Vlassov V. V, Leeflang M. M. G., Deeks J. J. (2009). Chapter 9: Assessing methodological quality BT - Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. In P. M. Deeks, J.J.; Gatsonis, C.; Bossuyt (Eds.), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.* Version 1.0.0. The Cochrane Collaboration. Available at: http://srdta.cochrane.org [Last accessed 03 June 2016].

Royston P., Altman D. G., Sauerbrei W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*, 25(1), 127-41. doi:10.1002/sim.2331.

Royston P., Moons K. G. M., Altman D. G., Vergouwe Y. (2009). Prognosis and prognostic research: Developing a prognostic model. *BMJ*, 338, b604. doi:10.1136/bmj.b604.

Steyerberg E. W., Moons K. G. M., van der Windt D. A., Hayden J. A., Perel P., Schroter S., Riley R. D., Hemingway H., Altman D. G. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med*, 10(2), e1001381. doi:10.1371/journal.pmed.1001381.

Taheriazam A., Sadatsafavi M., Moayyeri A. (2005). Outcome predictors in nonoperative management of newly diagnosed subacromial impingement syndrome: a longitudinal study. *MedGenMed*, 7(1), 63.

Tu Y.-K., Gilthorpe M. S. (2007). Revisiting the relation between change and initial value: a review and evaluation. *Stat Med*, 26(2), 443-57. doi:10.1002/sim.2538.

Vodicka E., Kim K., Devine E. B., Gnanasakthy A., Scoggins J. F., Patrick D. L. (2015). Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007-2013). *Contemp Clin Trials*, 43, 1-9. doi:10.1016/j.cct.2015.04.004.

Walker-Dilks C., Wilczynski N. L., Haynes R. B. (2008). Cumulative Index to Nursing and Allied Health Literature search strategies for identifying methodologically sound causation and prognosis studies. *Appl Nurs Res*, 21(2), 98-103. doi:10.1016/j.apnr.2006.04.004.

Wilczynski N. L., Haynes R. B. (2004). Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. *BMC Med.*, 2, 23. doi:10.1186/1741-7015-2-23.

Wilczynski N. L., Haynes R. B. (2005). Optimal search strategies for detecting clinically sound prognostic studies in EMBASE: an analytic survey. *J Am Med Inform Assoc*, 12(4), 481-5. doi:10.1197/jamia.M1752.

Wolff R., Whiting P., Moons K. (2015). *Systematic reviews of prognostic studies 1: assessing risk of bias in studies of prediction models using the PROBAST tool.* Workshop at 23rd Cochrane Colloquium, Vienna, 3-7 Oct 2015. Available at: http://2015.colloquium.cochrane.org/workshops/systematic-reviews-prognostic-studies-1-assessing-risk-bias-studies-prediction-models [Last accessed 03 June 2016].

Orientation Table Chapter 4

Part	Ch.	Title	Aims
ONE	1	General introduction, aims, content and structure of the thesis	1. To provide a general introduction to the topic 2. To summarise the aims, content and structure of the thesis
	2	Background	To provide the relevant topical and conceptual background to the programme of research
TWO	3	Prognostic models in adults undergoing physiotherapy for rotator cuff disorders - a systematic review	To establish the state of evidence on prognostic models in adults undergoing physiotherapy for painful rotator cuff disorders
	4	**Developing and validating the physiotherapy protocol for the prognostic study**	**1. To establish the state of evidence on the effectiveness of physiotherapy interventions for adults with painful atraumatic PTTs** **2. To develop and validate the physiotherapy treatment protocol**
	5	Selecting and defining the candidate prognostic factors for the prognostic study	1. To identify and select the candidate factors for the prognostic model study 2. To define the specific measures for the selected factors
	6	Predicting the outcome of physiotherapy in adults with painful partial-thickness rotator cuff tears (PTTs) – a prognostic model study	To develop a prognostic model for the outcome of a period of conservative treatment with physiotherapy in adult patients with painful atraumatic PTTs
	7	Drawing meaningful magnitude-based inferences from the prognostic study – Minimal Important Difference (MID) and responder analysis	1. To establish an estimate of the MID of the Western Ontario Rotator Cuff Index (WORC), the primary outcome of the prognostic model study 2. To apply the estimated MID to an exploratory responder analysis
THREE	8	Overall summary and conclusions	1. To summarise the research 2. To provide overall conclusions and consider implications
FOUR		Appendices	Appendices to Chapters 3-7

CHAPTER 4
Developing and validating the physiotherapy protocol for the prognostic study

4.1 BACKGROUND, AIMS AND OBJECTIVES

Physiotherapy is a standard conservative treatment for impingement-related shoulder complaints, including PTTs (see *Chapter 2*), and background searches of the research and descriptive literature on PTTs revealed that exercises and manual therapy are considered its principal components (Finnan & Crosby 2010, Fukuda 2003, Wolff et al. 2006). The prognostic study (*Chapter 6*) was designed to develop a prognostic model for the outcome of a phase of conservative treatment with physiotherapy in adults with painful atraumatic PTTs. I aimed to base the physiotherapy protocol for this study on the best evidence regarding exercises and manual therapy.

I had previously conducted a systematic review on "manual therapy and exercises for impingement-related shoulder pain" (Braun & Hanchard 2010). Although published prior to my PhD programme, Braun & Hanchard (2010) was thus highly relevant. However, during the development of the prognostic study protocol (May 2011 to May 2012), my routine searches revealed that several RCTs had been published since my 2010 review. I therefore conducted formal electronic searches (Cochrane Library, Medline, Embase, Cinahl, and PEDro) and hand-searches of the reference lists of relevant articles up to February 2012 to identify any systematic reviews postdating my own that might have incorporated some or all of these RCTs. These searches focused on systematic reviews explicitly addressing physiotherapy interventions including exercises or manual therapy, and identified five (Kelly et al. 2010, Littlewood et al. 2012, Dewhurst 2010, Kromer et al. 2009, Kuhn 2009). All addressed impingement-related shoulder pain, and all apparently included PTTs. Four of the reviews (Kelly et al. 2010, Littlewood et al. 2012, Dewhurst 2010, Kuhn 2009) explored the effectiveness of exercises, whereas one (Kromer et al. 2009) took a broader scope on physiotherapy interventions including exercises and manual therapy. Regarding these reviews' search cut-off dates, the most recent was November 2010 (Littlewood et al. 2012). Since I was aware of some RCTs that had been published subsequent to this, an update was indicated. I decided to build upon and update my own 2010 review (Braun & Hanchard 2010) in a further systematic review (Braun et al. 2013) which forms part of my PhD

programme of research and was also published in *Physical Therapy Reviews* (see further).

At issue was whether the most recent evidence would provide conclusive guidance on the optimal type, composition or dosage of exercises and manual therapy, which previous reviews, including my own (Braun & Hanchard 2010), could not (see further). The objectives were thus to identify and synthesise the evidence as to the effectiveness of manual therapy and exercises in general, and the most effective type, composition and dosage of manual therapy and exercise interventions in particular; and to document any reports of adverse events (such as exacerbations of symptoms, progression of PTTs).

This chapter summarises the two systematic reviews concerned and their implications for the prognostic study protocol. The emphasis is on the methods, results and agreement and disagreement with other systematic reviews. The reader is directed to *Appendix 4.1* for the published full report of the updated review (Braun et al. 2013). Because Braun et al. (2013) and Braun & Hanchard (2010) are complementary, they are aggregated as far as possible for the following summary which also draws out key points of interest. The review report (*sections 4.2* to *4.5*) is followed by an account of the development and validation of the physiotherapy protocol for the prognostic study (*sections 4.6* to *4.8*).

4.2 METHODS OF THE SYSTEMATIC REVIEWS

4.2.1 REVIEW DESIGN AND REPORTING

As Braun et al. (2013) was an explicit systematic review update, the methods were largely predefined by my previous review (Braun & Hanchard 2010). The reporting of Braun et al. (2013) followed the standards set out by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (Moher et al. 2010).

4.2.2 REVIEW CRITERIA

4.2.2.1 Types of research

Braun & Hanchard (2010) was a methodologically composite systematic review. It included systematic reviews of RCTs or quasi-RCTs, but also RCTs or quasi-RCTs that

post-dated the included reviews' search dates. Braun et al. (2013) specifically focussed on subsequent RCTs and quasi-RCTs. Both Braun & Hanchard (2010) and Braun et al. (2013) required reports to be in full, published form, in English or German.

4.2.2.2 Participants

The population of interest was defined as shoulder impingement rather than PTTs *per se*. There has been only minimal research specific to PTTs, but the clinical presentation of patients with painful PTTs is conventionally considered to be essentially that of shoulder impingement (see also *Chapter 2 section 2.5*), and shoulder impingement populations would be expected to largely incorporate people with painful PTTs. The population of interest was broadly operationalised as "patients with pain arising locally in [non-operated, non-traumatic] shoulders with grossly normal mobility" (Braun & Hanchard 2010 p. 63) in the presence of standard diagnostic criteria for shoulder impingement (Braun & Hanchard 2010; Braun et al. 2013). Trials exclusively considering FTTs were not included in Braun et al (2013), reflecting the more specific motivation for that review.

4.2.2.3 Interventions

Manual therapy was defined as any type of manual mobilisation or manipulation, which could be supplemented by therapeutic exercises. Exercises encompassed any type of active therapeutic regime, and could involve either supervised or home exercises, or both. Interventions combining manual therapy and exercises were acceptable. The interventions had to be delivered by physiotherapists. Additional physical therapy modalities, such as electrotherapy, were accepted only if they were supplementary to the manual therapy and exercise intervention. As comparator, any other treatment or no treatment was accepted.

4.2.2.4 Outcomes

The primary outcomes were pain, disability and self-perceived change of symptoms (Braun & Hanchard 2010; Braun et al. 2013), shoulder function (Braun & Hanchard 2010) and health-related quality of life (Braun et al. 2013). Further, adverse events were documented in both reviews.

4.2.3 SEARCHES AND SELECTION

Electronic searches were made of the Cochrane Library, Medline, Embase, Cinahl and PEDro and additionally, for Braun et al. (2013), two German databases, *Thieme Connect* and *Physiotherapeuten.de.* The searches were supplemented by hand searches of the reference lists of relevant articles. The cut-off date of the searches for the initial review was October 2008. The searches for the update covered the period to September 2012.

4.2.4 QUALITY ASSESSMENT, DATA EXTRACTION AND SYNTHESIS

The methodological quality of the included research was assessed with the AMSTAR checklist (systematic reviews) and the PEDro scale (RCTs and quasi-RCTs). Purpose-designed forms were used for the extraction of data on the key characteristics and results of the reviews and trials. Where possible, missing confidence intervals (CIs) were calculated using the Cochrane Collaboration's Review Manager (RevMan) software (version 5.1.7) (Cochrane Review Manager no date). To enhance interpretation of the findings' clinical relevance, estimates of the MID for any of the outcome measures, either as reported or imputed from the literature, were obtained and documented in the process of the update. Braun et al. (2013) further involved independent duplicate screening and assessment of methodological quality and data extraction by two reviewers. In cases of disagreement, consensus was sought through discussion, for which a third person was also available. All findings were analysed narratively. Meta-analysis was proposed where applicable.

4.3 RESULTS

4.3.1 CHARACTERISTICS OF THE INCLUDED RESEARCH

Braun & Hanchard (2010) identified eight systematic reviews and six subsequent RCTs as at October 2008, and Braun et al. (2013) identified nine further RCTs in the update period (see *Table 4.1* for the complete list). The systematic reviews were published between 2002 and 2006, with searches spanning 1966 to 2005. Five explicitly addressed impingement-related shoulder pain, one addressed rotator cuff tears and two addressed general shoulder pain but reported on subgroups of impingement-related shoulder pain (see *Table 4.1*). Two reviews (Desmeules et al. 2003, Trampas & Kitsios 2006) focussed on exercises and manual therapy, while the others had a

broader scope on interventions including exercises and manual therapy. The reviews covered any type of comparison such as no treatment, different types of manual therapy or exercises or surgery. The reviews were heterogeneous in terms of the definition of outcomes of interest, with two (Desmeules et al. 2003; Green & Alexander 2002) not defining any, and thus covered a variety of outcomes such as pain, disability functional limitations (including range of motion and strength), quality of life, sick leave, return to work, use of medication or adverse events (or side effects). The reviews included between five and 27 studies, only nine of which (represented by 10 reports, and totalling 562 participants) were RCTs or quasi-RCTs evaluating exercises or manual therapy.

The 15 RCTs (see *Table 4.1*), which were published between 2005 and 2012, totalled 1,072 participants. Thus, the participants in the systematic reviews and RCTs in combination totalled 1,634.

Table 4.1: Overview of systematic reviews and RCTs included in Braun & Hanchard (2010) and Braun et al. (2013)
Articles are ordered by publication type and year of publication.

Article ID	Publication type	Populations
Green 2002	SR	Shoulder pain
Johansson 2002	SR	Shoulder impingement
Desmeules 2003	SR	Shoulder impingement
Green 2003	SR	Shoulder pain
Ejnisman 2004	SR	Rotator cuff tears
Michener 2004	SR	Shoulder impingement
Faber 2006	SR	Shoulder impingement
Trampas 2006	SR	Shoulder impingement
Dickens 2005	RCT	Shoulder impingement
Giombini 2006	RCT	Shoulder impingement
Haahr 2006	RCT	Shoulder impingement
Senbursa 2007	RCT	Shoulder impingement
Cloke 2008	RCT	Shoulder impingement
Lombardi 2008	RCT	Shoulder impingement
Barbosa 2008	RCT	Shoulder impingement
Ketola 2009	RCT	Shoulder impingement
Engebretsen 2009, 2011	RCT	Shoulder impingement
Baskurt 2011	RCT	Shoulder impingement
Beaudreuil 2011	RCT	Shoulder impingement
Senbursa 2011	RCT	Shoulder impingement
Holmgren 2012	RCT	Shoulder impingement
Maenhout 2013*	RCT	Shoulder impingement
Subasi 2012	RCT	Shoulder impingement

*ID = first author, yr; SR = systematic review; *date of final print version - the report was available online 2012 before the systematic review cut date*

The diagnostic labels and criteria varied across the included reviews and RCTs, but all were compatible with a spectrum of impingement-related shoulder pain including PTTs. None of the included reviews or RCTs was specific to PTTs, however. The evidence primarily related to subacute (six weeks to three months) and chronic (more than three months) complaints; acute complaints were not represented. Implicitly, most of the RCTs were conducted in outpatient secondary care settings.

Numerous exercise or manual therapy interventions, or combinations of exercises and manual therapy, were investigated. Most of the RCTs or quasi-RCTs reviewed by others and us investigated exercise interventions. The interventions differed widely in terms of their overall duration; the number, frequency and duration of treatment sessions; the type, amount and composition of exercises and/or manual therapy techniques; and the intensity of exercises or grade of manual mobilisation. The spectrum of studied interventions included a variety of approaches to the strengthening of the rotator cuff, shoulder and scapular muscles; active and passive stretching of shoulder muscles and soft tissues; scapular positioning and stabilisation; humeral centring exercises; and active and passive (manual) mobilisation of the shoulder and shoulder region joints and soft tissues. In some studies, supplementary modalities such as cold packs or therapeutic ultrasound were applied. Each of the studied interventions was unique. Likewise, the interventions were compared with a variety of approaches such as waiting list controls; placebo treatment (e.g. sham ultrasound); a different type of exercises; the addition of manual therapy techniques to exercises; electrotherapeutic interventions (e.g. ultrasound); corticosteroid injections; or arthroscopic subacromial decompression surgery.

The most commonly assessed outcomes included pain, disability and shoulder function, and quality of life. These outcomes were variably defined and were assessed with a variety of outcome measures. Follow-up varied across the studies ranging from a few weeks to two and a half years.

Most of the systematic reviews synthesised their findings narratively. Only two reviews included meta-analyses of some results, but these did not relate to the outcomes of interest as defined for Braun & Hanchard (2010). The clinical heterogeneity of the supplementary trials made the synthesis of their findings difficult and, in combination with the heterogeneity of outcome measures, precluded meta-analysis.

4.3.2 METHODOLOGICAL QUALITY OF THE INCLUDED RESEARCH

The systematic reviews and the subsequent RCTs were of variable methodological quality as assessed by the AMSTAR checklist (Shea et al. 2007) and PEDro scale (Sherrington et al. 2000), respectively. All of the RCTs were affected by methodological limitations, i.e. carried some risk of bias.

4.3.3 EVIDENCE FOR THE EFFECTIVENESS OF EXERCISES AND MANUAL THERAPY

The following summary is presented separately for the systematic reviews and the subsequent trials to accommodate these studies' different levels of focus.

All of the systematic reviews in Braun & Hanchard (2010) concluded that some evidence (reportedly "weak" or "limited") supported the use of exercises; and all but one that there was support, though weak, for using exercises and manual therapy in combination. The exception was Ejnisman et al. (2004), who reported that the evidence was insufficient to enable a conclusion. None of the systematic reviews was able to establish optimal manual therapy or exercise parameters. None found sufficient evidence to conclude on the effectiveness of manual therapy as a stand-alone treatment. None reported on adverse events.

The RCTs included in Braun & Hanchard (2010) and Braun et al. (2013), which supplemented my review of reviews, supported those general consensus positions without establishing the additional certainty and detail required to justify development of a prescriptive treatment protocol (specifying the type, composition and dosage of exercises and manual therapy). Specifically, three RCTs compared exercises to no intervention or to different electrotherapies, of which two (Engebretsen et al. 2011, Lombardi et al. 2008), both well conducted, found evidence in favour of exercises for at least one important outcome; the third (Giombini et al. 2006) favoured electrotherapy but was at some risk of bias. One supplemental RCT (Barbosa et al. 2008) investigated the addition of manual mobilisation techniques to exercises and found that this significantly enhanced outcomes; however, the RCT was at high risk of bias. Three supplemental RCTs compared combinations of exercises and manual therapy techniques with or without adjunctive physical modalities to waiting list controls and found greater benefits in the intervention groups. Between group differences were statistically significant in two of these RCTs (Senbursa et al. 2007; Dickens et al. 2005), although Senbursa et al. (2007) was at high risk of bias and reported with internal inconsistencies; while the third RCT, by Cloke et al. (2008), was at some risk of bias. Two RCTs (Haahr & Andersen 2006, Ketola et al. 2009) compared exercise-based physiotherapy interventions with surgical decompression and found no significant differences. There were several head-to-head comparisons of different types of exercises. These were mostly inconclusive, although, based on well conducted single studies, "dynamic humeral centring" significantly improved pain at three months compared to nonspecific shoulder mobilisation exercises (Beaudreuil et al. 2011), and a specific strengthening strategy for the rotator cuff and scapular stabilisers conferred

significant benefit over a nonspecific home exercise programme on a number of patient-important outcomes (Holmgren et al. 2012). No supplemental RCT evaluated manual therapy as a stand-alone intervention.

Only three RCT reports covering two trials (Engebretsen et al. 2009 & 2011, Giombini et al. 2006) reported an a priori plan for collecting adverse events data and very few adverse events were reported. These related to exacerbations of pain. Meta-analysis of the RCTs was precluded by their clinical heterogeneity.

4.4 CONCLUSIONS

Braun & Hanchard (2010) and Braun et al. (2013) provided evidence from eight systematic reviews and 15 subsequent RCTs supporting, in principle, the use of exercises or exercises in combination with manual therapy in adults with subacute or chronic impingement-related shoulder pain, including PTT. Clinical heterogeneity and methodological deficiencies in the RCTs hindered synthesis and warranted cautious interpretation of the findings, such that it was not possible to conclude on the optimal type, composition or dosage of the interventions. In this regard, the update by Braun et al. (2013) failed to provide any conclusive advancement of the evidence as established by Braun & Hanchard (2010). In summary, while the body of available research evidence supports a combination of manual therapy and exercises for impingement-related shoulder pain, it provides no guidance on detail. Clinicians must therefore draw on their judgement to inform their choice of exercises, manual therapy techniques and treatment parameters.

Braun & Hanchard (2010) and Braun et al. (2013) revealed a need for further RCTs to establish the optimal type, composition and dosage of manual therapy and exercise interventions for impingement-related shoulder pain as a basis for a coherent body of evidence. Both reviews further noted a lack of studies on acute and rotator cuff tear (PTT) populations. Improvements in the quality of methods and reporting are essential.

4.5 AGREEMENT WITH OTHER SYSTEMATIC REVIEWS

The findings of my updated systematic review (Braun et al. 2013) overall largely agree with the further systematic reviews (November 2008 onwards) that I had identified from the literature (Kelly et al. 2010; Littlewood et al. 2012; Dewhurst 2010; Kromer et al. 2009; Kuhn 2009). Of these, only one (Kromer et al. 2009) conducted a meta-analysis on a relevant outcome, and then in a single instance. This was the pain outcome from

two trials evaluating the addition of manual therapy to exercises (standardized mean difference (95% CI): 0.88 (0.36-1.40)), which favoured the addition. The appropriateness of this pooling is questionable and the imprecision of the estimate perhaps unsurprising, considering the extent of clinical heterogeneity as well as the small sample size (n = 66). Indeed, Kromer et al. (2009) precluded pooling of other results, and Kelly et al. (2010) ruled out pooling any results on the grounds of clinical heterogeneity.

Two reviews (Dewhurst 2010, Kuhn 2009) presented an "evidence-based exercise protocol" derived from their syntheses. However, each of the protocols involved over-reaching from, or over-interpretation of, the evidence. Dewhurst (2010) provided a compilation of exercises with "an evidence-base from randomized controlled trials to improve symptoms of patients with subacromial impingement syndrome" (p. 112). The three systematic reviews and four RCTs from which the exercises were derived had been critically assessed and found to comply with some predefined quality criteria including "the standards of the Critical Appraisal Skills Programme (CASP)" (p. 112). However, five of these reports had been included in the reviews by Braun & Hanchard (2010) or Braun et al. (2013), and most had revealed methodological deficiencies warranting cautious interpretation. Beyond this, it seems that Dewhurst (2010) extracted any exercises that had been used in the studies' intervention groups without considering further aspects such as whether any of the exercises had been investigated by more than a single study. Kuhn (2009) presented a protocol "based on the best evidence demonstrating a beneficial effect for exercise in the treatment of rotator cuff tendonitis" (p. 156), derived from an arbitrary cut-off of "clinical significance" ($p < 0.05$ and difference of the effect size or difference between groups □ 20%). There seems to be no rational justification for this approach.

4.6 DRAFTING THE PHYSIOTHERAPY PROTOCOL

4.6.1 CONTENT OF PHYSIOTHERAPY

Braun & Hanchard (2010) and Braun et al. (2013) provided the basis for the treatment protocol for the prognostic model study (see *Chapter 6*). Without over-reaching from the inconclusive research evidence, there was no justification for defining any specific selection of exercises or manual therapy techniques which the collaborating physiotherapists would have to use in their treatment of study participants. Instead, the protocol rested upon the broad principles that:

□

1) exercises, preferably in combination with manual therapy, would be the key components of treatment; and
2) flexibility in the detail of the interventions as well as in the provision of adjunctive modalities would be allowed.

These principles lent themselves to a domain-based approach to the physiotherapy protocol, which was embodied in a study-specific physiotherapy report form (*Appendix 4.2*). In the form, implicit broad domains (e.g. exercises, mobilisations) were exploded into typical "indicative" intervention categories, e.g. "strengthening exercises focusing on the rotator cuff muscles", "stabilisation exercises", "manual mobilisation techniques (shoulder)" and "soft tissue techniques (shoulder or shoulder girdle)", which comprised the final categories. These final categories were informed by screening for and extracting "typical interventions" from Braun & Hanchard (2010) and Braun et al. (2013), initially in its unpublished form, the other systematic reviews identified from the literature (Kelly et al. 2010, Littlewood et al. 2012, Dewhurst 2010, Kromer et al. 2009, Kuhn 2009) and the descriptive literature on PTTs and impingement-related shoulder pain (see *Chapter 2 section 2.10*). On the report form, a tick box was provided adjacent to each category. This was to facilitate documentation of physiotherapy treatments. Individual categories were not intended to be prescriptive or mandatory, and the list was not intended to be exclusive (an "anything else" box was provided).

4.6.2 PROVISION OF PHYSIOTHERAPY

Although the content of sessions within the evidence-based framework was intended to be discretionary, the amount of physiotherapy and provision of adjunctive modalities and the maximum number of physiotherapy sessions could not be entirely so. This was not based on the research evidence, which was equivocal in both respects, but on the need for practitioners to comply with German healthcare regulations. As outlined in the background (*Chapter 2 section 2.10.4*), the standard maximum number of physiotherapy sessions for patients with a diagnosis of a PTT within the German statutory health insurance is 18 (three sets of of six sessions), with a recommended frequency of two or more sessions per week (GKV-Heilmittelrichtlinie 2011). Exceeding this limit may require the health insurer to approve a formal request. Adherence to the healthcare regulations is the responsibility of the treating doctor, and was in the context of the prognostic study thus the responsibility of Dr Betthäuser. Based on Dr Betthäuser's clinical experience and my own, we anticipated that the majority of the participants would require approximately 12 sessions (i.e. two prescriptions of six sessions), while some patients would require fewer, and some more. Further assuming

that the most common frequency of physiotherapy is one to two sessions per week, we anticipated that the length of observation of around three months would overall well fit within the statutory framework.

The duration of a single physiotherapy session in Germany is not strictly regulated, but the remuneration agreements between the German statutory health insurers and the German physiotherapy associations include recommended values for the different types of treatment. For standard physiotherapy or manual therapy, the current value is 15 to 25 minutes (vdek 2016). Many physiotherapy practices schedule a single session with 20 minutes (Gutefrage.net 2013). Some practices have longer sessions, and the common schedule of a single session for privately insured patients is 30 minutes (Privatpreise.de 2008). Thus, we expected the duration of single sessions of standard physiotherapy or manual therapy to mainly range from 20 to 30 minutes.

4.7 PILOTING THE REPORT FORM

I invited six colleagues to pilot the form and to provide feedback. All were practicing clinicians and three held Master degrees. Thus, clinical and academic perspectives were represented. I asked them to consider the content of the form, i.e. whether they thought it covered all relevant domains, or whether anything relevant was missing or redundant, as well as its usability, i.e. whether they found it quick and easy to use and complete. I asked them whether they would suggest any changes, either to the content or to the design of the form. All six colleagues piloted the form and provided feedback. Regarding the content, all found that it covered most of the domains that they considered relevant. Three colleagues suggested the addition of “spinal mobilisation” as a further domain, because they regularly applied manual mobilisations to the cervical or thoracic spine to their patients with impingement-related shoulder pain. Although the use of cervical or thoracic mobilisations for improving clinical outcomes in patients with impingement-related shoulder pain was not backed by evidence and hardly mentioned in the literature at that time, I considered it reasonable to allow for the addition of spinal mobilisations or other contingencies, and did so by adding a line titled “anything else”. Regarding the usability of the form, all six colleagues found that it was very easy and quick to complete, and that they did not experience any problems with completing it. Thus, no further changes were necessary.

4.8 VALIDATION BY COLLABORATING PHYSIOTHERAPY PRACTICES

I initially approached, by telephone, leads from a convenience sample of eleven practices over the wider Hamburg area with a view to recruitment (further details are provided in *Chapter 6*), but also with a view to determining whether the protocol reflected their usual approach to the treatment of patients with impingement-related shoulder pain. Of these, seven were eligible and agreed to collaborate, and the leads of these practices confirmed that the protocol, and the domain structure of the physiotherapy report form, complied with their usual approach to the treatment of patients with impingement-related shoulder pain (with or without PTTs). This endorsed my strategy.

4.9 SUMMARY

The physiotherapy protocol for the prognostic model study as part of this PhD programme of research was based on the available evidence on exercises and manual therapy for adults with impingement-related shoulder pain (including PTTs) and took into account the relevant German healthcare regulations. The level of detail of the available evidence only lent itself to a broad domain-based protocol, which, though, fitted perfectly with clinical practice. Piloting and feedback from targeted practice endorsed the protocol and the implicit domain structure of the physiotherapy report form.

4.10 REFERENCES

Barbosa R. I., Goes R., Mazzer N., Fonseca M. C. R. (2008). The influence of joint mobilization on tendinopathy of the biceps brachii and supraspinatus muscles. *Braz J Phys Ther*, *12*(4), 298-303.

Baskurt Z., Ba kurt F., Gelecek N., Özkan M. H. (2011). The effectiveness of scapular stabilization exercise in the patients with subacromial impingement syndrome. *J Back Musculoskelet Rehabil*, 24(3), 173-9. doi:10.3233/BMR-2011-0291.

Beaudreuil J., Lasbleiz S., Richette P., Seguin G., Rastel C., Aout M., Vicaut E., Cohen-Solal M., Lioté F., de Vernejoul M.-C., Bardin T., Orcel P. (2011). Assessment of dynamic humeral centering in shoulder pain with impingement syndrome: a randomised clinical trial. *Ann Rheum Dis*, 70(9), 1613-8. doi:10.1136/ard.2010.147694.

Braun C., Hanchard N. C. A. (2010). Manual therapy and exercise for impingementrelated shoulder pain. *Phys Ther Rev*, 15(2), 62-83. doi:10.1179/174328810X12786297204675.

Braun C., Bularczyk M., Heintsch J., Hanchard N. C. A. (2013). Manual therapy and exercises for shoulder impingement revisited. *Phys Ther Rev*, 18(4), 263-84. doi:10.1179/108331913X13709388114510.

Cloke D. J., Watson H., Purdy S., Steen I. N., Williams J. R. (2008). A pilot randomized, controlled trial of treatment for painful arc of the shoulder. *J Shoulder Elbow Surg*, 17(1 Suppl), 17S-21S. doi:10.1016/j.jse.2007.07.005.

Cochrane Review Manager (no date). Most recent version available at: http://tech.cochrane.org/revman [Last accessed 10 June 2016].

Desmeules F., Côté C. H., Frémont P. (2003). Therapeutic exercise and orthopedic manual therapy for impingement syndrome: a systematic review. *Clin J Sport Med*, 13(3), 176-82.

Dewhurst A. (2010). An exploration of evidence-based exercises for shoulder impingement syndrome. *Int Musculoskelet Med*, 32(3), 111-6.

Dickens V., Williams J., Bhamra M. (2005). Role of physiotherapy in the treatment of subacromial impingement syndrome: a prospective study. *Physiotherapy*, 91(3), 159-64. doi:10.1016/j.physio.2004.10.008.

Ejnisman B., Andreoli C. V, Soares B. G. O., Fallopa F., Peccin M. S., Abdalla R. J., Cohen M. (2004). Interventions for tears of the rotator cuff in adults. *Cochrane Database Syst Rev*, (1), CD002758.

Engebretsen K., Grotle M., Bautz-Holter E., Ekeberg O. M., Juel N. G., Brox J. I. (2011). Supervised exercises compared with radial extracorporeal shock-wave therapy for subacromial shoulder pain: 1-year results of a single-blind randomized controlled trial. *Phys Ther*, 91(1), 37-47. doi:10.2522/ptj.20090338.

Engebretsen K., Grotle M., Bautz-Holter E., Sandvik L., Juel N. G., Ekeberg O. M., Brox J. I. (2009). Radial extracorporeal shockwave treatment compared with supervised exercises in patients with subacromial pain syndrome: single blind randomised study. *BMJ*, 339, b3360. doi:10.1136/bmj.b3360.

Faber E., Kuiper J. I., Burdorf A., Miedema H. S., Verhaar J. A. N. (2006). Treatment of Impingement Syndrome: A Systematic Review of the Effects on Functional Limitations and Return to Work. *J Occup Rehabil*, 16(1), 6-24. doi:10.1007/s10926-005-9003-2.

Finnan R. P., Crosby L. A. (2010). Partial-thickness rotator cuff tears. *J Shoulder Elbow Surg*, 19(4), 609-16. doi:10.1016/j.jse.2009.10.017.

Fukuda H. (2003). The management of partial-thickness tears of the rotator cuff. *J Bone Joint Surg Br*, 85(1), 3-11. doi:10.1302/0301-620X.85B1.13846.

Giombini A., Di Cesare A., Safran M. R., Ciatti R., Maffulli N. (2006). Short-term effectiveness of hyperthermia for supraspinatus tendinopathy in athletes: a short-term randomized controlled study. *Am J Sports Med*, 34(8), 1247-53. doi:10.1177/0363546506287827.

GKV-Heilmittelrichtlinie (2011). *Heilmittelrichtlinie und -katalog [Healthcare directive and catalogue]* (der gesetzlichen Krankenversicherung, GKV [of the German statutory health insurance, GKV]). Available at: https://www.gkv-spitzenverband.de/krankenversicherung/ambulante_leistungen/heilmittel/heilmittelrichtlinie/heilmittel-richtlinie.jsp [Last accessed 03 June 2016].

Green S., Alexander M. (2002). Evidence-based clinical statement: physiotherapy interventions for shoulder pain. *Aust. Physiother. Assoc.* Available at: http://www.ibrarian.net/navon/paper/Australian_Physiotherapy_Association_Shoulder_Pai.pdf?paperid=9742111 [Last accessed 03 June 2016].

Green S., Buchbinder R., Hetrick S. (2003). Physiotherapy interventions for shoulder pain. *Cochrane Database Syst Rev*, (2), CD004258. doi:10.1002/14651858.CD004258.

Gutefrage.net. (2013). *Physiotherapie im 20-Minuten-Takt?* [*Physiotherapy in 20 minute rhythm?]* Available at: http://www.gutefrage.net/frage/physiotherapie-im-20min-takt [Last accessed 03 June 2016].

Haahr J. P., Ostergaard S., Dalsgaard J., Norup K., Frost P., Lausen S., Holm E. A., Andersen J. H. (2005). Exercises versus arthroscopic decompression in patients with subacromial impingement: a randomised, controlled study in 90 cases with a one year follow up. *Ann Rheum Dis*, 64(5), 760-4. doi:10.1136/ard.2004.021188.

Haahr J. P., Andersen J. H. (2006). Exercises may be as efficient as subacromial decompression in patients with subacromial stage II impingement: 4-8-years' follow-up in a prospective, randomized study. *Scand J Rheumatol*, 35(3), 224-8. doi:10.1080/03009740600556167.

Holmgren T., Björnsson Hallgren H., Öberg B., Adolfsson L., Johansson K. (2012). Effect of specific exercise strategy on need for surgery in patients with subacromial impingement syndrome: randomised controlled study. *BMJ*, 344, e787. doi:10.1136/bmj.e787.

Johansson K., Oberg B., Adolfsson L., Foldevi M. (2002). A combination of systematic review and clinicians' beliefs in interventions for subacromial pain. *Br J Gen Pract*, 52(475), 145-52.

Kelly S. M., Wrightson P. A., Meads C. A. (2010). Clinical outcomes of exercise in the management of subacromial impingement syndrome: a systematic review. *Clin Rehabil*, 24(2), 99-109. doi:10.1177/0269215509342336.

Ketola S., Lehtinen J., Arnala I., Nissinen M., Westenius H., Sintonen H., Aronen P., Konttinen Y. T., Malmivaara A., Rousi T. (2009). Does arthroscopic acromioplasty provide any additional value in the treatment of shoulder impingement syndrome?: a two-year randomised controlled trial. *J Bone Joint Surg Br*, 91(10), 1326-34. doi:10.1302/0301-620X.91B10.22094.

Kromer T. O., Tautenhahn U. G., de Bie R. A., Staal J. B., Bastiaenen C. H. G. (2009). Effects of physiotherapy in patients with shoulder impingement syndrome: a systematic review of the literature. *J Rehabil Med*, 41(11), 870-80. doi:10.2340/16501977-0453.

Kuhn J. E. (2009). Exercise in the treatment of rotator cuff impingement: a systematic review and a synthesized evidence-based rehabilitation protocol. *J Shoulder Elbow Surg*, 18(1), 138-60. doi:10.1016/j.jse.2008.06.004.

Littlewood C., Ashton J., Chance-Larsen K., May S., Sturrock B. (2012). Exercise for rotator cuff tendinopathy: a systematic review. *Physiotherapy*, 98(2), 101-9. doi:10.1016/j.physio.2011.08.002.

Lombardi I., Magri A. G., Fleury A. M., Silva A. C. Da, Natour J. (2008). Progressive resistance training in patients with shoulder impingement syndrome: a randomized controlled trial. *Arthritis Rheum*, 59(5), 615-22.

Maenhout A. G., Mahieu N. N., De Muynck M., De Wilde L. F., Cools A. M. (2013). Does adding heavy load eccentric training to rehabilitation of patients with unilateral subacromial impingement result in better outcome? A randomized, clinical trial. *Knee Surg Sports Traumatol Arthrosc*, 21(5), 1158-67. doi:10.1007/s00167-012-2012-8.

Michener L. A., Walsworth M. K., Burnet E. N. (2004). Effectiveness of rehabilitation for patients with subacromial impingement syndrome: a systematic review. *J Hand Ther*, 17(2), 152-64.

Moher D., Liberati A., Tetzlaff J., Altman D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg*, 8(5), 336-41. doi:10.1016/j.ijsu.2010.02.007.

Privatpreise.de (2008). *Gebührenübersicht für Therapeuten: Therapiedauer [Overview of fees for therapists: duration of treatment].* Available at: http://www.privatpreise.de/privatpreise/therapiedauer.html [Last accessed 03 June 2016].

Senbursa G., Baltaci G., Atay A. (2007). Comparison of conservative treatment with and without manual physical therapy for patients with shoulder impingement syndrome: a prospective, randomized clinical trial. *Knee Surg Sports Traumatol Arthrosc*, 15(7), 915-21. doi:10.1007/s00167-007-0288-x.

Senbursa G., Baltaci G., Atay Ö. A. (2011). The effectiveness of manual therapy in supraspinatus tendinopathy. *Acta Orthop Traumatol Turc*, 45(3), 162-7. doi:10.3944/AOTT.2011.2385.

Shea B. J., Grimshaw J. M., Wells G. A., Boers M., Andersson N., Hamel C., Porter A. C., Tugwell P., Moher D., Bouter L. M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*, 7, 10. doi:10.1186/1471-2288-7-10.

Sherrington C., Herbert R. D., Maher C. G., Moseley A. M. (2000). PEDro. A database of randomized trials and systematic reviews in physiotherapy. *Man Ther*, 5(4), 223-6. doi:10.1054/math.2000.0372.

Subasi, V.; Toktas, H.; Demirdal, Ü.S.; Türel, A.; Cakir, T.; Kavuncu V. (2012). Water-based versus land-based exercise program for the management of shoulder impingement syndrome. *Turk J Phys Med Rehab*, 58(2), 79-84.

Trampas A., Kitsios A. (2006). Exercise and manual therapy for the treatment of impingement syndrome of the shoulder: a systematic review. *Phys Ther Rev*, 11(2), 125-42.

vdek (2016). *Vergütungsvereinbarung gemäß §125 SGB V für die Abrechnung physiotherapeutischer Leistungen, Massage und medizinischer Bäder (West) [Remuneration agreement for physiotherapy interventions, massage and medicinal baths]* (of the Verband der Ersatzkassen [Association of Health Insurance Companies], vdek. Available via: https://www.kvhb.de/sites/default/files/ek-preisliste-physio.pdf [Last accessed 26 June 2016].

Wolff A. B., Sethi P., Sutton K. M., Covey A. S., Magit D. P., Medvecky M. (2006). Partial-thickness rotator cuff tears. *J Am Acad Orthop Surg*, 14(13), 715-25.

Orientation Table Chapter 5

Part	Ch.	Title	Aims
ONE	1	General introduction, aims, content and structure of the thesis	1. To provide a general introduction to the topic 2. To summarise the aims, content and structure of the thesis
	2	Background	To provide the relevant topical and conceptual background to the programme of research
TWO	3	Prognostic models in adults undergoing physiotherapy for rotator cuff disorders - a systematic review	To establish the state of evidence on prognostic models in adults undergoing physiotherapy for painful rotator cuff disorders
	4	Developing and validating the physiotherapy protocol for the prognostic study	1. To establish the state of evidence on the effectiveness of physiotherapy interventions for adults with painful atraumatic PTTs 2. To develop and validate the physiotherapy treatment protocol
	5	**Selecting and defining the candidate prognostic factors for the prognostic study**	**1. To identify and select the candidate factors for the prognostic model study** **2. To define the specific measures for the selected factors**
	6	Predicting the outcome of physiotherapy in adults with painful partial-thickness rotator cuff tears (PTTs) – a prognostic model study	To develop a prognostic model for the outcome of a period of conservative treatment with physiotherapy in adult patients with painful atraumatic PTTs
	7	Drawing meaningful magnitude-based inferences from the prognostic study – Minimal Important Difference (MID) and responder analysis	1. To establish an estimate of the MID of the Western Ontario Rotator Cuff Index (WORC), the primary outcome of the prognostic model study 2. To apply the estimated MID to an exploratory responder analysis
THREE	8	Overall summary and conclusions	1. To summarise the research 2. To provide overall conclusions and consider implications
FOUR		Appendices	Appendices to Chapters 3-7

CHAPTER 5
Selecting and defining the candidate prognostic factors for the prognostic study

5.1 INTRODUCTION

Central to the design of my prognostic model study (*Chapter 6*) was the selection of a set of candidate prognostic factors. Numerous factors were potentially relevant, meaning that choices had to be made. The number of candidate factors to be investigated in the prognostic study was limited to 10 at the outset (as explained in *Chapter 6*). In this chapter, I describe the selection process for the candidate factors, including the specific measures and/or measurement systems for their use in the prognostic study.

5.2 AIMS AND OBJECTIVES

The overarching aim was to maximise the chances of developing a valid, high-performing prognostic model by optimising the selection of the candidate factors. To achieve this, a three-stage approach comprising identification, critical assessment and final selection of prognostic factors was taken.

5.3 METHODS

5.3.1 SELECTION APPROACH

The three-stage selection approach is outlined in *Figure 5.1*.

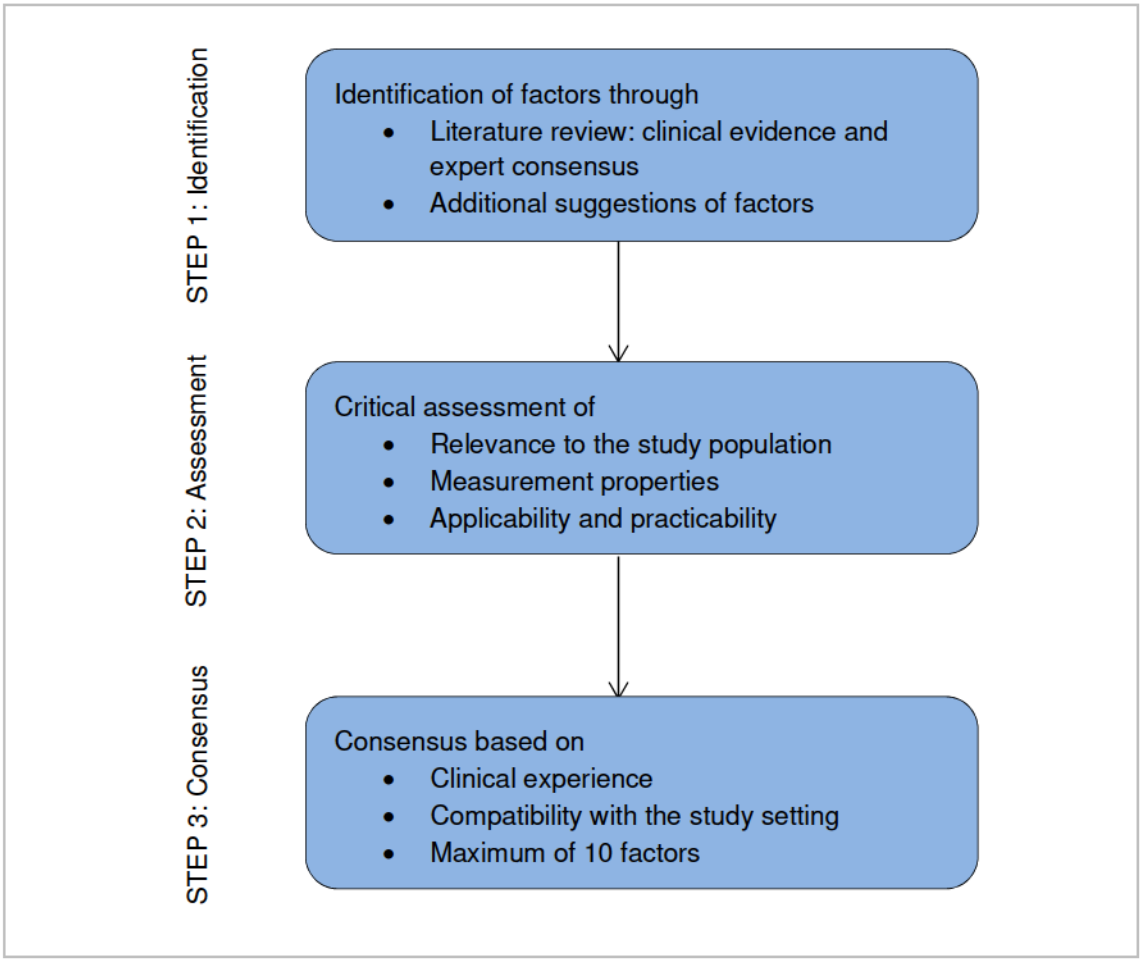


Figure 5.1: Identification and selection of candidate factors – outline of process

5.3.2 STAGE 1 - IDENTIFICATION OF POTENTIALLY RELEVANT FACTORS

5.3.2.1 Literature review

A literature review was conducted, concluding in January 2012. This followed a rigorous systematic approach which included the *a priori* setting out of inclusion and selection criteria. Given my prior knowledge of the scarcity and limitations of prognosis research on PTTs and rotator cuff disorders, I chose a broad approach to the identification of factors from the literature.

Criteria for considering studies

- *Study types*

  Any primary longitudinal studies investigating potential associations between baseline factors and clinical outcomes, or systematic reviews of such studies, were

considered. No restrictions were made on the study design or analysis methods. Further, I considered studies such as Delphi studies or surveys which aimed to establish expert consensus on potentially relevant prognostic factors. Inclusion was restricted to publications in English or German. There was no restriction on the publication type.

- *Participants*

  The review focussed on populations of adults with any type of painful rotator cuff disorder. Evidence from general shoulder pain populations was also considered for systematic reviews and expert consensus studies.

- *Interventions*

  Inclusion of prognostic studies was limited to studies in which participants were followed over a course of conservative treatment. Studies in which not all participants received conservative treatment, or where this was unclear, were not considered. There was no restriction on the type or duration of treatment.

- *Prognostic factors*

  Prognostic factors were only considered if they were assessable at the baseline assessment. No restriction was made on the type of factors (e.g. demographic, physical, psychological).

- *Outcomes*

  Any outcome was considered.

- *Data sources and searches*

  Searches were made of Medline, Embase, Cinahl, Cochrane Library and PEDro databases. These were supplemented by hand searches of the reference lists in relevant articles. The primary search strategy is shown in *Appendix 5.1*. The terms and component strings (population, interventions, prognosis) were used in all possible permutations. All searches were conducted by me, and up to February 2012.

- *Study selection and data extraction*

  All publications that met the above criteria were included. From each included study (or review of studies), I extracted any prognostic factors available at baseline that were considered by the study investigators (or review authors) to have a relevant association with a study outcome. I placed no restriction on the outcome other than it needed to be clinically relevant.

5.3.2.2 Additional suggestions of candidate factors

When I started to plan the prognostic study, I had in mind several factors that I considered of potential interest. These arose from my clinical experience, discussion with clinicians and knowledge of the literature on rotator cuff disorders. Among these factors were age, disability and symptom duration, which I knew had some support of their prognostic relevance from clinical evidence.

Two other factors that I considered as potentially relevant were diabetes and smoking. The detrimental effects of both these on many aspects of health are well-known (Alberg 2008, Hoogwerf et al. 2006). I was also aware that both diabetes and smoking had been linked with the prevalence of various musculoskeletal complaints including shoulder pain (Gauri & Fatima 2009, Palmer et al. 2003, Porter & Hanley 2001), and I found there was some evidence for a higher prevalence of rotator cuff disorders in smokers (Baumgarten et al. 2010, Viikari-Juntura et al. 2008) and in people with diabetes (Miranda et al. 2005, Ranger et al. 2015, Viikari-Juntura et al. 2008).

Both diabetes and smoking had been shown to impair the healing of musculoskeletal tissues such as bone (Gaston & Simpson 2007). And while I did not find specific evidence in relation to tendon healing, this seemed plausible. Although I found no evidence relating to the role of diabetes and smoking in predicting clinical outcomes in people with rotator cuff disorders, there was limited evidence for inferior functional outcomes after surgical repair of rotator cuff tears in smokers (Mallon et al. 2004) and in people with diabetes (Clement et al. 2010). I thus concluded that there was a strong rationale for including both smoking and diabetes as candidate factors and put these forward for inclusion in the list of putative prognostic factors, regardless of whether or not they were identified through the searches.

All factors that were identified from the literature search and the additionally suggested factors were taken together; these formed the initial compilation of potential candidate factors.

5.3.3 STAGE 2 - CRITICAL ASSESSMENT OF IDENTIFIED FACTORS

Each of the above factors was critically assessed by me for its compliance with the following criteria:

- *Relevance to my study:* the factor had to be relevant to the population and setting of the prognostic study.

- *Measurement properties*: the factor had to be assessable with a sufficiently valid and reliable measurement. In the case of PROMS, this included the availability of validated German versions.
- *Applicability/practicability* in clinical practice: the factor had to be applicable in most clinical settings, i.e. should not require any special equipment, diagnostic skills or techniques beyond those commonly available and covered by standard statutory health care. Also, the factor had to be applicable within Dr Betthäuser's routine clinical practice.

Following their assessment, the factors that met all three criteria were then placed into two broad groups. (The criteria were derived in the knowledge of the availability of the expert consensus study.)

1) Group A: These were factors with reasonably consistent support for their prognostic relevance, either clinical evidence from three or more studies, or from both clinical evidence *and* expert consensus.
2) Group B: These were factors with either limited support for their relevance from clinical evidence but no support from expert consensus; factors with no support from clinical research but support from expert consensus; or any additionally suggested factors for which there was no support from clinical evidence or expert consensus.

5.3.4 STAGE 3 - CONSENSUS ON FINAL TEN FACTORS

All factors that had been retained at the second screening stage were discussed with Dr Betthäuser and Dr Hanchard to achieve consensus on the 10 factors to be included in the prognostic study.

The findings from the selection process were documented and summarised by text and tabulation.

5.4 RESULTS

5.4.1 STAGE 1 - IDENTIFICATION OF FACTORS

5.4.1.1 Literature review

Clinical evidence

After screening overall around 3,900 records from the searches and checking reference lists, I identified 23 primary study reports for 22 studies that met the pre-specified inclusion criteria. The study reports (same as any other identified reports) and their references are listed in *Appendix 5.2*. For completeness, the references are also included in the reference list of this chapter. Only two of these studies, one published in full (Yamanaka & Matsumoto 1994) and the other published in a conference abstract only (Selvanetti et al. 1998) exclusively involved patients with PTTs. I also identified one systematic review of primary prognostic studies on populations of patients with various shoulder complaints (Kuijpers et al. 2004). The overlap between the systematic review and my compilation of primary studies was two studies (Brox & Brevik 1996, Morrison et al. 1997).

I extracted 34 factors from the primary studies. The systematic review did not provide any factors additional to those reported in the primary studies. Diabetes and smoking were not among these 34 factors. The evidence base for the majority of the factors consisted only of one or two clinical studies. The factors were very heterogeneous, as were the approaches used for their measurement. For only three factors - namely, age, disability and symptom duration - was there reasonably consistent evidence of prognostic value from several (five or more) studies pertaining to clinical outcomes of conservative treatment in patients with rotator cuff disorders. The systematic review (Kuijpers et al. 2004) provided some evidence in support of the prognostic relevance of pain, age, duration of symptoms and baseline disability in painful shoulder disorders.

Expert consensus

I found one study which involved a Delphi expert consensus approach on the “prediction of persistent shoulder pain in general practice” (Vergouw et al. 2011). Through a three-round Delphi process, a multidisciplinary panel of health professionals (general practitioners, orthopaedists, physiotherapists and manual therapists), all of

whom were “involved in or having thorough knowledge of shoulder pain in clinical practice” (Methods para.1), selected and consented a set of 10 factors which were considered as most important to predict the persistence of shoulder pain in the general practice setting. These factors were: symptom duration, history of symptoms/shoulder pain, severity of shoulder disability, age, shoulder pain intensity, coexisting neck pain, multisite pain, fear-avoidance beliefs, illness perceptions and pain catastrophizing (Table 2). The last five of these were factors which were not identified from the primary studies.

5.4.1.2 Additionally suggested factors

As planned, diabetes and smoking were retained for further consideration despite not being identified via the literature review.

5.4.1.3 Initial compilation of potentially relevant factors

In summary, 36 factors formed the initial compilation. They are presented in *Table 5.1*. The factors are grouped according to various categories:

- Demographic factors
- Activity-related factors
- Symptom-related factors
- Factors related to history of symptoms/shoulder pain
- Factors from physical examination
- Factors related to comorbidities and (self-reported) health status
- Psychosocial/psychological factors
- Structural factors
- Rotator-cuff specific factors
- Interventional factors
- Economical factors

Table 5.1 also shows the sources of support for each factor (i.e. clinical evidence and/or expert consensus).

Table 5.1: Compilation of factors, grouped by type, showing source of supporting evidence and results of the selection
See foot of table for key

No	Factor	Clinical evidence (article IDs)	Expert consensus	Stage 2 out-come	Stage 3 out-come
Demographic factors					
1	Age	Ekeberg 2010; Kennedy 2006a,b; Maman 2009; Morrison 1997; Selvanetti 1998; Yamanaka 1994; SR Kuijpers 2004	E	▯ A	▯
2	Education	Engebretsen 2010		▯ B	▯
3	Sex	Ekeberg 2010; Kennedy 2006a,b		▯ B	▯
Activity-related factors					
4	Dominant arm (affected)	Chard 1988		▯ B	▯
5	Patient's physical demands (overhead athletes)	Selvanetti 1998		▯ B	▯
Symptom-related factors					
6	Coexisting neck pain		E	▯ B	▯
7	Disability (shoulder-related)	Bartolozzi 1994; Ekeberg 2010; Engebretsen 2010; Hung 2010; Kennedy 2006a,b; Taheriazam 2005; SR Kuijpers 2004	E	▯ A	▯
8	Pain	Ekeberg 2010; Kennedy 2006a,b; SR Kuijpers 2004	E	▯ A	▯
9	Patient's global rating of (severity of) problem	Kennedy 2006a,b		▯ B	▯
10	Quality of sleep (sleep loss due to shoulder pain)	Hawkins 1995		▯ B	▯

▯

No	Factor	Clinical evidence (article IDs)	Expert consesus	Stage 2 out-come	Stage 3 out-come
Factors related to history of symptoms/shoulder pain					
11	Aetiology of symptoms (including history of trauma to the shoulder, type of rotator cuff disease and overuse)	Chard 1988; Selvanetti 1998; Yamanaka 1994		□	
12	History of (previous) shoulder pain	Engebretsen 2010	E	□ A	□
13	Symptom duration	Bartolozzi 1994; Chard 1988; Kennedy 2006a,b; Morrison 1997; Taheriazam 2005; SR Kuijpers 2004	E	□ A	□
Factors from physical examination					
14	Impingement sign (presence/absence)	Tanaka 2010		□	
15	Muscle strength (serratus anterior; abduction; rotation)	Hung 2010; Itoi & Tabata 1992; Vad 2002		□	
16	Range of motion (active: abduction, external rotation; passive)	Ekeberg 2010; Itoi & Tabata; Minagawa 2003; Tanaka 2010; Vad 2002		□	
17	Scapular kinematics: internal rotation	Hung 2010		□	
Factors related to comorbidities and (self-reported) health status					
18	Diabetes	None direct; see text (section 5.3.2.2)		□ B	□
19	Glenohumeral arthritis	Vad 2002		□	
20	Health status (self reported)	Engebretsen 2010; Kennedy 2006a,b		□ B	□
21	Multisite pain		E	□	
22	Smoking	None direct; see text (section 5.3.2.2)		□ B	□
Psychosocial/psychological factors					
23	Fear-avoidance beliefs		E	□	

□

No	Factor	Clinical evidence (article IDs)	Expert consesus	Stage 2 out-come	Stage 3 out-come
24	Illness perceptions		E	□ B	□
25	Pain catastrophizing		E	□ B	□
Structural factors (shoulder)					
26	Acromion type/morphology	Morrison 1997; Taheriazam 2005; Wang 2000		□	
27	Humeral head migration	Vad 2002		□	
28	Osseous abnormalities (not further specified)	Selvanetti 1998		□	
Rotator cuff specific factors					
29	Fatty infiltration	Maman 2009		□	
30	Muscle atrophy	Tanaka 2010; Vad 2002		□	
31	Tear size (extent)	Selvanetti 1998; Yamanaka 1994		□	
32	Type of rotator cuff pathology or tear; tendon integrity	Bartolozzi 1994; Maman 2009; Tanaka 2010		□	
Interventional factors					
33	Corticosteroid injections (response to initial injection; previous)	Cummins 2009; Ekeberg 2010		□	
34	Medication (regular medication; over-the-counter medication)	Ekeberg 2010, Kennedy 2006a		□	
Economical factors					
35	Insurance (worker's compensation) claims	Hawkins 1995, Kennedy 2006a,b		□	
36	Sick leave	Ekeberg 2010		□	

*Article ID = first author, yr; E = expert consensus; □ = included (retained); □ **A** = reasonably consistent support for prognostic relevance (see section 5.3.3); **B** = limited support for prognostic relevance (see 5.3.3); □ □ = excluded; SR = systematic review.*

□

5.4.2 STAGE 2 - CRITICAL ASSESSMENT

The results of the screening at stage 2 are presented in the column “stage 2” in *Table 5.1*. Nineteen factors were excluded at this step; these are listed together with the main reason for exclusion for each factor in *Appendix 5.3*.

As described above, the remaining 17 factors were placed into one of two groups (A or B) according to the extent of the supporting evidence (see *section 5.3.3*). Five factors (age, disability, pain, duration of symptoms and history of shoulder pain) were supported by evidence (group A), whereas the remaining factors were not (group B) (see *Table 5.1*).

5.4.3 STAGE 3 – FINAL SELECTION STAGE FOR THE TEN CANDIDATE FACTORS

Discussions with Dr Betthäuser and Dr Hanchard resulted in the final selection of 10 candidate factors. These included all five factors in group A. The results of the consensus process are presented in the column “stage 3” in *Table 5.1*. The 10 candidate factors for inclusion in the prognostic study are separately presented in *Table 5.2*.

Table 5.2: Final selection of candidate prognostic factors

No	Factor	Category / type
1	Age	Demographic
2	Sex	Demographic
3	Physical demands	Activity-related
4	Disability	Symptom-related
5	Pain	Symptom-related
6	History of shoulder pain	History of symptoms
7	Symptom duration	History of symptoms
8	Diabetes	Comorbidities
9	Smoking	Comorbidities
10	Pain catastrophizing	Psychological

5.5 SELECTION AND RATIONALE FOR THE SPECIFIC MEASURES AND MEASUREMENT OF THE CANDIDATE PROGNOSTIC FACTORS

The selection of the 10 candidate prognostic factors was followed by the careful consideration of specific measures and measurement systems for each factor for use in the prognostic study. Although Dr Betthäuser assessed all of the factors as part of his usual practice, these were not usually measured in a suitable way for the purposes of the prognostic study. The heterogeneity in the measures (e.g. PROMS to measure disability) and measurement systems (e.g. different cut-offs for age) used in the primary clinical studies meant these were of very limited use as a guide.

Table 5.3 provides an overview of the measure/measurement system used for each of the 10 candidate factors in the prognostic study. The factors were mostly assessed through a self-report questionnaire (this is further explained in the report of the prognostic study, see *Chapter 6*). With the exception of self-explanatory demographic factors (i.e. age and sex), the rationale for the selection of the measure/measurement system for each factor is detailed below.

Table 5.3: The candidate factors and their measures used in the study

No	Predictor variable	Measure / measurement system
1	Age	Age at initial presentation (yr)
2	Sex	Sex (female, male)
3	Physical demands	”Before you had your current shoulder problem, did a typical week include one or more of the following activities: • Repetitive or prolonged use of the affected arm for strength effort (e.g. lifting, carrying or moving heavy loads, athletic sports, strength-demanding skilled manual work) • Repetitive or prolonged use of the arm above shoulder height (e.g. overhead work, overhead sports, throwing sports, work as a hairdresser)?” (yes/no)
4	Disability	Western Ontario Rotator Cuff Index (WORC) (Kirkley et al. 2003a); validated German version (Huber et al. 2005) (score)
5	Pain	“What is the worst amount of pain that you have experienced within the past week?” (100 mm visual analogue scale VAS)
6	History of shoulder pain (incl. previous treatment)	“Prior to the current episode, have you ever seen a medical doctor or therapist for pain in this shoulder?” (yes, no)
7	Symptom duration	“For how long have you been having your current shoulder complaints?” (weeks)
8	Diabetes	“Do you have diabetes?” (yes/no)
9	Smoking	“Are you a smoker? Please tick “yes” if you regularly smoke at least once a week any amount of tobacco” (yes/no)
10	Pain catastrophizing	Pain Catastrophizing Scale (PCS; Sullivan et al. 1995); validated German version (Meyer et al. 2008) (score)

5.5.1 ACTIVITY-RELATED FACTORS

5.5.1.1 Physical demands

The measurement of this factor was based on clinical knowledge; these are typical questions when taking a patient history. The intention was to capture any of the key activities that typically provoke or increase pain in patients with impingement-related

shoulder pain. The question was pre-tested by Dr Betthäuser in a small sample of patients to ensure sufficient comprehensibility. No problems were encountered.

5.5.2 SYMPTOM-RELATED FACTORS

5.5.2.1 Disability

This factor was measured by use of a PROM. In order to select a suitable measure for use in the prognostic study I searched the literature for PROMS assessing shoulder disability (with or without inclusion of additional aspects such as HRQoL). The aim was to find, if available, a rotator cuff specific PROM that ideally satisfied several pre-defined criteria which are presented in *Table 5.4*. I intended to use the selected PROM both for the assessment of disability as a prognostic factor and as the primary outcome of the prognostic study.

Table 5.4: Criteria for selecting a disability PROM

No	Criteria
1	Is the PROM rotator-cuff specific, or, if not, does it appear suitable for use in the study population?
2	Does it assess disability (with or without additional aspects such as pain or HRQoL)?
3	Does it have satisfactory psychometric properties?
4	Is a validated German version available?
5	Is the PROM easy for patients to understand and complete (availability and quality of instructions)?
6	Is it sufficiently quick to complete?
7	Does it appear to be well accepted by patients (any data on problems)?
8	Is it available free of charge?
9	Is it sufficiently easy and quick to analyse?

To gain a comprehensive overview of available instruments, I focussed on reviews of any shoulder disability measures, and identified three (Bot et al. 2004, Kirkley et al. 2003b, Michener & Leggin 2001, Wright & Baumgarten 2010). A book on “classifications and scores of the shoulder” (Habermeyer et al. 2006) as well as several further more focussed publications on shoulder disability measures (Desai et al. 2010,

Ekeberg et al. 2008, Michener & Leggin 2001, Oh et al. 2009, Razmjou et al. 2006, Roy & Esculier 2011, Roy et al. 2009) served as additional sources of information.

I extracted all reported PROMs from the publications. I identified 26, which can be viewed in *Appendix 5.4*. I categorised them into “general” (further sub-grouped into either “upper extremity-related” or “shoulder-related” measures) and “specific” (further sub-grouped into “condition-specific” and “disease-specific”) instruments. Of the 26 PROMS, only four were disease-specific and only two had been developed specifically for use in people with rotator cuff disorders (see also *Chapter 2 section 2.9.1*): the Western Ontario Rotator Cuff Index (WORC) (Huber et al. 2005, Kirkley et al. 2003a) and the Rotator Cuff Quality of Life measure (RC_QOL) (Hollinshead et al. 2000).

Both the WORC and the RC_QOL appeared suitable for consideration. After initially inspecting the questionnaires I decided to further inspect the WORC for compatibility with the pre-specified criteria to determine its suitability for use in my study, the primary reason being that the RC_QOL appeared to be very long (34 items compared to 21 items of the WORC), meaning that it would take longer for patients to complete. The results can be viewed in *Table 5.5* and show that the WORC was well suited for use in the prognostic study.

Table 5.5: Assessing the suitability of the WORC

No	Criteria	□ / □	Details
1	Is the PROM rotator cuff specific, or, if not, does it appear suitable for use in the study population?	□	The WORC was specifically designed for use in people with rotator cuff pathologies (including tendinitis, tears (PTT and FTT) and cuff arthropathy) (Kirkley et al. 2003a).
2	Does it assess disability (with or without additional aspects such as pain or HRQoL)?	□	The WORC assesses disability in combination with pain and QoL.
3	Does it have satisfactory psychometric properties?	□	The WORC has been shown to be a valid, reliable and responsive PROM for use in people with rotator cuff disorders (Huber et al. 2005, Kirkley et al. 2003a,b, Roy & Esculier 2011, Wright & Baumgarten 2010).
4	Is a validated German version available?	□	Huber et al. (2005)
5	Is the PROM easy for patients to understand and complete (availability and quality of instructions)?	□	The WORC comes along with very helpful instructions on each item.
6	Is it sufficiently quick to complete?	□	On average 7.5 min as reported by Huber et al. (2005)
7	Does it appear to be well-accepted by patients (any data on problems)?	□	There is some evidence of good acceptance, i.e. no problems with completing the WORC (Huber et al. 2005, Kirkley et al. 2003a).
8	Is it available free of charge?	□	
9	Is it sufficiently easy and quick to analyse?	□	Analysis requires a bit more effort due to the use of VAS scales, but appears acceptable.

□ *= satisfactorily fulfilled;* □ *= not satisfactorily fulfilled*

□

The WORC (see *Appendix 5.5*) consists of 21 questions (items), which are subdivided into five domains: “physical symptoms” (items 1-6), “sports/recreation” (items 7-10 items), “work” (items 11-14), “lifestyle” (items 15-18) and “emotions” (items 19-21). Each question is answered by putting a mark on a 100 mm visual analogue (VAS) scales (with the right end indicating the highest extent of symptoms or disability), which leads to a score between 0 (no symptoms no disability) and 2100 (greatest extent of symptoms and disability). The score can alternatively be presented as percentage of normal by applying the following formula: (2100 – total score) / 2100 x 100 (Kirkley et al. 2003a p. 87).

5.5.2.2 Pain

Baseline pain intensity was assessed with a VAS. VAS are widely used to measure pain intensity in clinical populations and are quick and easy to administer (Hawker et al. 2011, Kamper 2012). VAS scales are considered a reliable and responsive measurement showing a high correlation with numerical rating scales (NRS) and other measures of pain (Hawker et al. 2011). VAS have been claimed to show some disadvantages compared to NRS, which include that they may be more difficult to understand, especially for older people (Hjermstad et al. 2011, Schomacher 2008), and that patients appear to prefer NRS over VAS (Hjermstad et al. 2011). The key reason for choosing a VAS for use in the study was that the WORC items (some of which relate to pain) are similarly rated using VAS scales. This minimised the number of different concepts of measurement in the questionnaires that the patients needed to become familiar with, thereby potentially reducing patient confusion.

The anchor for the assessment of the baseline pain intensity (“What is the worst amount of pain that you have experienced within the past week?”) was chosen based on my clinical experience of the varying intensity of pain, i.e. its dependency of pain to shoulder movements. A 100 mm long VAS was chosen, with the verbal descriptor “no pain” at the left end and “worst imaginable pain” at the right end. These are reportedly the most widely used descriptors used in research on VAS (Hjermstad et al. 2011).

5.5.3 FACTORS RELATED TO HISTORY OF SYMPTOMS/SHOULDER PAIN

5.5.3.1 History of shoulder pain

Linking the patients' history of shoulder pain to the consultation of a healthcare professional (doctor ot therapist) enabled the assessment of both previous episodes of and previous treatment (by any healthcare professional) for pain in the affected shoulder. Considering only episodes that led to the consultation of a healthcare professional further introduced a threshold of severity. This question was pre-tested in clinical practice on a small sample of patients by Dr Betthäuser to ensure that it was sufficiently comprehensible. No problems were encountered.

5.5.4 FACTORS RELATED TO COMORBIDITIES AND (SELF-REPORTED) HEALTH STATUS

5.5.4.1 Diabetes

I decided on a simple binary measure ("yes" or "no") for diabetes. This appeared reasonable given the lack of direct evidence on the role of diabetes in painful rotator cuff disorders and prevalence data. Since the estimated prevalence of (diagnosed) diabetes in the German population was approximately 7% at the time when the prognostic study was planned (Diabetes Deutschland 2012), any further categorisation of diabetes (such as by type of diabetes or treatment) would have introduced a threat to obtaining appropriate numbers of cases for the statistical analysis.

5.5.4.2 Smoking

The consideration of a threshold of dose for smoking, although arbitrarily chosen, was included to reflect the accumulating evidence on the increasing prevalence and health risks of light or intermittent smoking (Schane et al. 2010). As light smokers may not consider themselves as smokers, it has been suggested to use more focused questions to also detect light smokers rather than just asking "Are you a smoker?" (Schane et al. 2009). In view of the limited knowledge on the role of smoking in rotator cuff disorders, it seemed reasonable to keep the question broad. The estimated overall prevalence of smokers in the German population (including both daily and intermittent smokers) was around 30% in 2011 (Statista.com_Rauchen no date). Thus, similar as

with diabetes, any further categorisation would have introduced a threat to obtaining appropriate numbers of cases for the statistical analyses.

5.5.5 PSYCHOLOGICAL/PSYCHOSOCIAL FACTORS

5.5.5.1 Pain catastrophizing

Pain catastrophizing has been defined as “an exaggerated negative ‘mental set’ brought to bear during actual or anticipated painful experience” (Sullivan et al. 2001 p. 53). It refers to the interpretation of pain as being “extremely threatening” (Leeuw et al. 2007 p. 79).

An investigation of the literature for available measures of pain catastrophizing yielded the Pain Catastrophizing Scale (PCS) (Sullivan et al. 1995). This validated, generic 13-item patient-reported instrument designed for both clinical and non-clinical populations appeared to be the only available PROM explicitly designed to assess pain catastrophizing. I checked it for suitability for its use in my prognostic study and also against the same criteria as the WORC (*Table 5.5*). The results presented in *Table 5.6* show that the PCS appeared well suited for use in the prognostic study.

The PCS instrument can be viewed in *Appendix 5.6*. The PCS consists of 13 items which represent statements describing different thoughts and emotions that may be associated with pain. Patients are asked to rate the degree to which they have each of these thoughts and emotions when they are experiencing pain by a five-point scale ranging from 0 (“not at all”) to 4 (“all the time”). The total PCS score can thus range from 0 to 52, with 0 representing no pain catastrophizing and 52 representing the highest possible degree of pain catastrophizing. Sullivan (2009) has proposed a PCS score of 30 as the “cut-off score for clinically relevant levels of catastrophizing” (p.7).

Table 5.6: Assessing the suitability of the PCS

No	Criteria	□ / □	Details
1	Is the PROM rotator cuff specific, or, if not, does it appear suitable for use in the study population?	□	The PCS is a generic measure that has been developed for use in various clinical and non-clinical populations. As such, it was appropriate for use in the prognostic study population.
2	Does it assess pain catastrophizing?	□	It was explicitly designed to assess pain catastrophizing (Sullivan et al. 1995).
3	Does it have satisfactory psychometric properties?	□	The PCS has been shown to be a sufficiently valid and reliable instrument (Meyer et al. 2008; Osman et al. 2000; Osman et al. 1997; Sullivan et al. 1995).
4	Is a validated German version available?	□	Meyer et al. (2008)
5	Is the PROM easy for patients to understand and complete (availability and quality of instructions)?	□	Appears to be; the introductory instructions on how to complete the PCS appear easy to understand.
6	Is it sufficiently quick to complete?	□	< 5 minutes (Sullivan 2009)
7	Does it appear to be well-accepted by patients (any data on problems)?	□	There is some evidence of good acceptance, i.e. no relevant problems with completing the PCS (Meyer et al. 2008).
8	Is it available free of charge?	□	Available via: http://sullivan-painresearch.mcgill.ca/pcs.php [last accessed 11 June 2016]
9	Is it sufficiently easy and quick to analyse?	□	Very easy and quick to analyse by summing up the scores for all items (Sullivan et al. 1995)

□ *= yes;* □ *= no*

□

5.6 SUMMARY

The three-stage process for selecting the 10 candidate factors for the prognostic model study comprised a systematic literature review to identify prognostic factors reported in clinical studies and other sources; a screening process to select factors which were relevant to the study, had suitable measurement properties, and were applicable and practical in clinical practice; and a consensus process which included an appraisal of the evidence base for each factor. Overall 36 factors were identified, most of which were evidenced from single studies. After screening, 19 factors were excluded. From the remaining 17 factors, 10 were chosen for inclusion in the study through a consensus process. Careful consideration was given to the selection of the specific measures and/or measurement systems for each factor for their use in the prognostic model study.

5.7 REFERENCES

Alberg A. J. (2008). Cigarette smoking: health effects and control strategies. *Drugs Today (Barc)*, 44(12), 895-904. doi:10.1358/dot.2008.44.12.1308898.

Bartolozzi A., Andreychik D., Ahmad S. (1994). Determinants of outcome in the treatment of rotator cuff disease. *Clin Orthop Relat Res*, (308), 90-7.

Baumgarten K. M., Gerlach D., Galatz L. M., Teefey S. a, Middleton W. D., Ditsios K., Yamaguchi K. (2010). Cigarette smoking increases the risk for rotator cuff tears. *Clin Orthop Relat Res*, 468(6), 1534-41. doi:10.1007/s11999-009-0781-2.

Bot S. D. M., Terwee C. B., van der Windt D. A. W. M., Bouter L. M., Dekker J., de Vet H. C. W. (2004). Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis*, 63(4), 335-41. doi:10.1136/ard.2003.007724.

Brox J. I., Brevik J. I. (1996). Prognostic factors in patients with rotator tendinosis (stage II impingement syndrome) of the shoulder. *Scand J Prim Health Care*, 14(2), 100-5.

Chard M. D., Sattelle L. M., Hazleman B. L. (1988). The long-term outcome of rotator cuff tendinitis--a review study. *Br J Rheumatol*, 27(5), 385-9.

Clement N. D., Hallett A., Macdonald D., Howie C., McBirnie J. (2010). Does diabetes affect outcome after arthroscopic repair of the rotator cuff? *J Bone Joint Surg Br*, 92(8), 1112-7. doi:10.1302/0301-620X.92B8.23571.

Conroy D. E., Hayes K. W. (1998). The effect of joint mobilization as a component of comprehensive treatment for primary shoulder impingement syndrome. *J Orthop Sports Phys Ther*, 28(1), 3-14. doi:10.2519/jospt.1998.28.1.3.

Cummins C. A., Sasso L. M., Nicholson D. (2009). Impingement syndrome: Temporal outcomes of nonoperative treatment. *J Shoulder Elb Surg*, 18(2), 172-7. doi:10.1016/j.jse.2008.09.005.

Desai A. S., Dramis A., Hearnden A. J. (2010). Critical appraisal of subjective outcome measures used in assessment of shoulder disability. *Ann R Coll Surg Engl*, 92(1), 9-13. doi:10.1308/003588410X12518836440522.

Diabetes Deutschland. (2012). *Prävalenz Diabetes (Deutscher Gesundheitssurvey) [Prevalence of diabetes (German health survey].* Available at: *http://www.diabetes-deutschland.de/aktuellesituation.html* [Last accessed 03 June 2016].

Ekeberg O. M., Bautz-Holter E., Tveitå E. K., Keller A., Juel N. G., Brox J. I. (2008). Agreement, reliability and validity in 3 shoulder questionnaires in patients with rotator cuff disease. *BMC Musculoskelet Disord*, 9, 68. doi:10.1186/1471-2474-9-68.

Ekeberg O. M., Bautz-Holter E., Juel N. G., Engebretsen K., Kvalheim S., Brox J. I. (2010). Clinical, socio-demographic and radiological predictors of short-term outcome in rotator cuff disease. *BMC Musculoskelet Disord*, 11(1), 239. doi:10.1186/1471-2474-11-239.

Engebretsen K., Grotle M., Bautz-Holter E., Ekeberg O. M., Brox J. I. (2010). Predictors of shoulder pain and disability index (SPADI) and work status after 1 year in patients with subacromial shoulder pain. *BMC Musculoskelet Disord*, 11, 218. doi:10.1186/1471-2474-11-218.

Gaston M. S., Simpson A. H. R. W. (2007). Inhibition of fracture healing. *J Bone Joint Surg Br*, 89(12), 1553-60. doi:10.1302/0301-620X.89B12.19671.

Gauri L. A., Fatima Q. (2009). Musculoskeletal manifestations of diabetes mellitus. *J Indian Med Assoc*, 107(11), 810, 819-21.

Habermeyer P., Magosch P., Lichtenberg S. (2006). *Classifications and Scores of the Shoulder*. Heidelberg: Springer.

Hardy D. C., Vogler J. B., White R. H. (1986). The shoulder impingement syndrome: prevalence of radiographic findings and correlation with response to therapy. , *AJR Am J Roentgenol,* 147(3), 557-61. doi:10.2214/ajr.147.3.557.

Hawker G. A., Mian S., Kendzerska T., French M. (2011). Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res (Hoboken),* 63(S11), S240-52. doi:10.1002/acr.20543.

Hawkins R. H., Dunlop R. (1995). Nonoperative treatment of rotator cuff tears. *Clin Orthop Relat Res*, (321), 178-88.

Hjermstad M. J., Fayers P. M., Haugen D. F., Caraceni A., Hanks G. W., Loge J. H., Fainsinger R., Aass N., Kaasa S. (2011). Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. *J Pain Symptom Manage*, 41(6), 1073-93. doi:10.1016/j.jpainsymman.2010.08.016.

Hollinshead R. M., Mohtadi N. G., Vande Guchte R. a, Wadey V. M. (2000). Two 6-year follow-up studies of large and massive rotator cuff tears: comparison of outcome measures. *J Shoulder Elbow Surg*, *9*(5), 373-81. doi:10.1067/mse.2000.108389.

Hoogwerf B. J., Sferra J., Donley B. G. (2006). Diabetes mellitus--overview. *Foot Ankle Clin*, 11(4), 703-15. doi:10.1016/j.fcl.2006.06.014.

Huber W., Hofstaetter J. G., Hanslik-Schnabel B., Posch M., Wurnig C. (2005). [Translation and psychometric testing of the Western Ontario Rotator Cuff Index (WORC) for use in Germany]. Z Orthop Ihre Grenzgeb, 143(4), 453-60. doi:10.1055/s-2005-836677.

Hung C.-J., Jan M.-H., Lin Y.-F., Wang T.-Q., Lin J.-J. (2010). Scapular kinematics and impairment features for classifying patients with subacromial impingement syndrome. *Man Ther*, 15(6), 547-51. doi:10.1016/j.math.2010.06.003.

Itoi E., Tabata S. (1992). Conservative treatment of rotator cuff tears. *Clin Orthop Relat Res*, (275), 165-73.

Kamper S. J. (2012). Pain intensity ratings. *J Physiother*, 58(1), 61. doi:10.1016/S1836-9553(12)70077-2.

Kennedy C. A., Haines T., Beaton D. E. (2006a). Eight predictive factors associated with response patterns during physiotherapy for soft tissue shoulder disorders were identified. *J Clin Epidemiol*, 59(5), 485-96. doi:10.1016/j.jclinepi.2005.09.003.

Kennedy C. A., Manno M., Hogg-Johnson S., Haines T., Hurley L., McKenzie D., Beaton D. E. (2006b). Prognosis in soft tissue disorders of the shoulder: predicting both change in disability and level of disability after treatment. *Phys Ther*, 86(7), 1013-32; discussion 1033–7.

Kirkley A., Alvarez C., Griffin S. (2003a). The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. *Clin J Sport Med*, 13(2), 84-92.

Kirkley A., Griffin S., Dainty K. (2003b). Scoring systems for the functional assessment of the shoulder. *Arthroscopy*, 19(10), 1109-20. doi:10.1016/j.arthro.2003.10.030.

Kuijpers T., van Der Windt D. A. W. M., van der Heijden G. J. M. G., Bouter L. M. (2004). Systematic review of prognostic cohort studies on shoulder disorders. *Pain*, 109(3), 420-31. doi:10.1016/j.pain.2004.02.017.

Leeuw M., Goossens M. E. J. B., Linton S. J., Crombez G., Boersma K., Vlaeyen J. W. S. (2007). The Fear-Avoidance Model of Musculoskeletal Pain: Current State of Scientific Evidence. *J Behav Med*, 30(1), 77-94. doi:10.1007/s10865-006-9085-0.

Mallon W. J., Misamore G., Snead D. S., Denton P. (2004). The impact of preoperative smoking habits on the results of rotator cuff repair. *J Shoulder Elb Surg*, 13(2), 129-32. doi:10.1016/j.jse.2003.11.002.

Maman E., Harris C., White L., Tomlinson G., Shashank M., Boynton E. (2009). Outcome of nonoperative treatment of symptomatic rotator cuff tears monitored by magnetic resonance imaging. *J Bone Joint Surg Am*, 91(8), 1898-906. doi:10.2106/JBJS.G.01335.

Meyer K., Sprott H., Mannion A. F. (2008). Cross-cultural adaptation, reliability, and validity of the German version of the Pain Catastrophizing Scale. *J Psychosom Res*, 64(5), 469-78. doi:10.1016/j.jpsychores.2007.12.004.

Michener L. A., Leggin B. G. (2001). A review of self-report scales for the assessment of functional limitation and disability of the shoulder. *J Hand Ther*, 14(2), 68-76.

Miranda H., Viikari-Juntura E., Heistaro S., Heliövaara M., Riihimäki H. (2005). A population study on differences in the determinants of a specific shoulder disorder versus nonspecific shoulder pain without clinical findings. *Am J Epidemiol*, 161(9), 847-55. doi:10.1093/aje/kwi112.

Morrison D. S., Frogameni A. D., Woodworth P. (1997). Non-operative treatment of subacromial impingement syndrome. *J Bone Joint Surg Am*, 79(5), 732-7.

Oh J. H., Jo K. H., Kim W. S., Gong H. S., Han S. G., Kim Y. H. (2009). Comparative evaluation of the measurement properties of various shoulder outcome instruments. *Am J Sports Med*, 37(6), 1161-8. doi:10.1177/0363546508330135.

Osman A., Barrios F. X., Kopper B. A., Hauptmann W., Jones J., O'Neill E. (1997). Factor structure, reliability, and validity of the Pain Catastrophizing Scale. *J Behav Med*, 20(6), 589–605.

Osman A., Barrios F. X., Gutierrez P. M., Kopper B. A., Merrifield T., Grittmann L. (2000). The pain catastrophizing scale: Further psychometric evaluation with adult samples. *J Behav Med*, 23(4), 351–365. doi:10.1023/A:1005548801037.

Palmer K. T., Syddall H., Cooper C., Coggon D. (2003). Smoking and musculoskeletal disorders: findings from a British national survey. *Ann Rheum Dis*, 62(1), 33-6.

Porter S. E., Hanley E. N. (2001). The musculoskeletal effects of smoking. *J Am Acad Orthop Surg*, 9(1), 9-17.

Ranger T. A., Wong A. M. Y., Cook J. L., Gaida J. E. (2015). Is there an association between tendinopathy and diabetes mellitus? A systematic review with meta-analysis. *Br J Sports Med*, Nov 23 [Epub ahead of print]. doi:10.1136/bjsports-2015-094735.

Razmjou H., Bean A., van Osnabrugge V., MacDermid J. C., Holtby R. (2006). Cross-sectional and longitudinal construct validity of two rotator cuff disease-specific outcome measures. *BMC Musculoskelet Disord*, 7, 26. doi:10.1186/1471-2474-7-26.

Roy J.-S., MacDermid J. C., Woodhouse L. J. (2009). Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum*, 61(5), 623-32. doi:10.1002/art.24396.

Roy J.-S., Esculier J.-F. (2011). Psychometric evidence for clinical outcome measures assessing shoulder disorders. *Phys Ther Rev*, 16(5), 331-46. doi:10.1179/1743288X11Y.0000000043.

Safran O., Schroeder J., Bloom R., Weil Y., Milgrom C. (2011). Natural history of nonoperatively treated symptomatic rotator cuff tears in patients 60 years old or younger. *Am J Sports Med*, 39(4), 710-4. doi:10.1177/0363546510393944.

Schane R. E., Glantz S. A., Ling P. M. (2009). Social smoking implications for public health, clinical practice, and intervention research. *Am J Prev Med*, 37(2), 124-31. doi:10.1016/j.amepre.2009.03.020.

Schane R. E., Ling P. M., Glantz S. A. (2010). Health effects of light and intermittent smoking: a review. *Circulation*, 121(13), 1518-22. doi:10.1161/CIRCULATIONAHA.109.904235.

Schomacher J. (2008). Gütekriterien der visuellen Analogskala zur Schmerzbewertung (Quality critaria of the visual analogue scale for pain assessment). *Physioscience*, 4(3), 125-33.

Selvanetti A., Giombini A., Caruso I. (1998). Nonoperative treatment of partial-thickness rotator cuff tears in overhead athletes. *Med Sci Sports Exerc*, 30(5), S260.

Statista.com_Rauchen (no date). Prävalenz Rauchen [prevalence of smoking] 2011. Available at: http://de.statista.com/statistik/daten/studie/261015/umfrage/praevalenz-des-rauchens-in-deutschland-nach-geschlecht/ [Last accessed 03 June 2016].

Sullivan M. J. L., Bishop S. R., Pivik J. (1995). The Pain Catastrophizing Scale: Development and validation. *Psychol Assess*, 7(4), 524-32. doi:10.1037//1040-3590.7.4.524.

Sullivan M. J., Thorn B., Haythornthwaite J. A., Keefe F., Martin M., Bradley L. A., Lefebvre J. C. (2001). Theoretical perspectives on the relation between catastrophizing and pain. *Clin J Pain*, 17(1), 52-64.

Sullivan M. J. L. (2009). *Pain Catastrophizing Scale. User Manual*. Available at: http://sullivan-painresearch.mcgill.ca/pdf/pcs/PCSManual_English.pdf [Last accessed 03 June 2016].

Taheriazam A., Sadatsafavi M., Moayyeri A. (2005). Outcome predictors in nonoperative management of newly diagnosed subacromial impingement syndrome: a longitudinal study. *MedGenMed*, 7(1), 63.

Tanaka M., Itoi E., Sato K., Hamada J., Hitachi S., Tojo Y., Honda M., Tabata S. (2010). Factors related to successful outcome of conservative treatment for rotator cuff tears. *Ups J Med Sci*, 115(3), 193-200. doi:10.3109/03009734.2010.493246.

Vad V. B., Warren R. F., Altchek D. W., O'Brien S. J., Rose H. A., Wickiewicz T. L. (2002). Negative prognostic factors in managing massive rotator cuff tears. *Clin J Sport Med*, 12(3), 151-7.

Vergouw D., Heymans M. W., de Vet H. C., van der Windt D. A., van der Horst H. E. (2011). Prediction of persistent shoulder pain in general practice: Comparing clinical consensus from a Delphi procedure with a statistical scoring system. *BMC Fam Pract*, 12(1), 63. doi:10.1186/1471-2296-12-63.

Viikari-Juntura E., Shiri R., Solovieva S., Karppinen J., Leino-Arjas P., Varonen H., Kalso E., Ukkola O. (2008). Risk factors of atherosclerosis and shoulder pain--is there an association? A systematic review. *Eur J Pain*, 12(4), 412-26. doi:10.1016/j.ejpain.2007.08.006.

Virta L., Mortensen M., Eriksson R., Möller M. (2009). How many patients with subacromial impingement syndrome recover with physiotherapy? A follow-up study of a supervised exercise programme. *Adv Physiother*, 11(3), 166-73. doi:10.1080/14038190802460481.

Wang J. C., Horner G., Brown E. D., Shapiro M. S. (2000). The relationship between acromial morphology and conservative treatment of patients with impingement syndrome. *Orthopedics*, 23(6), 557-9.

Wright R. W., Baumgarten K. M. (2010). Shoulder outcomes measures. *J Am Acad Orthop Surg*, 18(7), 436-44.

Wu H. P., Dubinsky T. J., Richardson M. L. (2003). Association of shoulder sonographic findings with subsequent surgical treatment for rotator cuff injury. *J Ultrasound Med*, 22(2), 155-61.

Yamanaka K., Matsumoto T. (1994). The joint side tear of the rotator cuff. A followup study by arthrography. *Clin Orthop Relat Res*, (304), 68-73.

Orientation Table Chapter 6

Part	Ch.	Title	Aims
ONE	1	General introduction, aims, content and structure of the thesis	1. To provide a general introduction to the topic 2. To summarise the aims, content and structure of the thesis
	2	Background	To provide the relevant topical and conceptual background to the programme of research
TWO	3	Prognostic models in adults undergoing physiotherapy for rotator cuff disorders - a systematic review	To establish the state of evidence on prognostic models in adults undergoing physiotherapy for painful rotator cuff disorders
	4	Developing and validating the physiotherapy protocol for the prognostic study	1. To establish the state of evidence on the effectiveness of physiotherapy interventions for adults with painful atraumatic PTTs 2. To develop and validate the physiotherapy treatment protocol
	5	Selecting and defining the candidate prognostic factors for the prognostic study	1. To identify and select the candidate factors for the prognostic model study 2. To define the specific measures for the selected factors
	6	**Predicting the outcome of physiotherapy in adults with painful partial-thickness rotator cuff tears (PTTs) – a prognostic model study**	**To develop a prognostic model for the outcome of a period of conservative treatment with physiotherapy in adult patients with painful atraumatic PTTs**
	7	Drawing meaningful magnitude-based inferences from the prognostic study – Minimal Important Difference (MID) and responder analysis	1. To establish an estimate of the MID of the Western Ontario Rotator Cuff Index (WORC), the primary outcome of the prognostic model study 2. To apply the estimated MID to an exploratory responder analysis
THREE	8	Overall summary and conclusions	1. To summarise the research 2. To provide overall conclusions and consider implications
FOUR		Appendices	Appendices to Chapters 3-7

CHAPTER 6
Predicting the outcome of physiotherapy in adults with painful partial-thickness rotator cuff tears (PPTs) – a prognostic model study

6.1 INTRODUCTION

The background literature review for this programme of research (see *Chapter 2*) and the systematic review of prognostic models (see *Chapter 3*) showed considerable uncertainty regarding indications for any treatment of people with painful PTTs, but especially conservative treatment. Conservative treatment is the first line for patients with PTTs and commonly includes physiotherapy, with possible adjuncts such as oral pain medication or subacromial corticosteroid injections. Despite this, no valid and usable prognostic model was available for the outcome of conservative treatment with physiotherapy for any type of rotator cuff disease, and no prospective prognostic model study had specifically addressed a population with painful PTTs. These findings highlighted the need for well-designed prognostic modelling research into the outcome of conservative treatment with physiotherapy in patients with painful PTTs. The present chapter reports on such a prognostic model study, which I designed and conducted as a key element of my PhD programme. The originality of this work is evidenced by the findings from the prognosis systematic review, which showed that at the time my study was planned, and indeed to date (June 2016)[5], there were no completed or registered on-going prospective prognostic model studies addressing the same research question.

6.2 STUDY AIMS AND PRIMARY RESEARCH QUESTION

The primary aim was to develop a prognostic model for the outcome of a phase of conservative treatment with physiotherapy in adults with painful PTTs. The primary research question was thus: what combination of factors can best predict the outcome of a period of conservative treatment with physiotherapy (with or without adjunctive medical treatment) in adults with shoulder pain and ultrasonographically diagnosed, atraumatic PTTs?

[5] Last update of database searches (systematic review strategy): 14 June 2016

Secondary aims were to determine the incidence of PTT progression and to establish participants' perceived global change in their shoulder complaints over the observation period. The latter was to enable an estimate of the MID of the WORC, the primary outcome of this study. The MID analysis, in itself a novel contribution to knowledge, is reported in *Chapter 7*.

6.3 ETHICS AND GOVERNANCE COMMITTEE APPROVALS

Approval for this study and the MID analysis was required from two ethics committees: the Teesside University School of Health and Social Care (SHSC) Research Governance & Ethics Committee (RGEC), and the Ethics Commission of the Hamburg Medical Council (the responsible German local commission). SHSC RGEC approval was obtained on 23 May 2012 (*Appendix 6.1*). This conferred cover under the University's Public Liability and Professional Indemnity scheme on the condition that all collaborating physiotherapy practices would provide written evidence of indemnity. Approval by the Ethics Commission of the Hamburg Medical Council was obtained on 09 November 2012 [6] (*Appendix 6.2*). The ethics approvals applied to both the prognostic study and the MID analysis.

6.4 STUDY PROTOCOL AND REGISTRATION

The study was developed from an *a priori* protocol (*Appendix 6.3*), which, in turn, was based on the information required by the ethics applications. It was retrospectively registered in the German Clinical Trials Register (Deutsches Register Klinischer Studien, DRKS[7]), which feeds into ICTRP, the International Clinical Trials Registry Platform of the WHO[8], on 08 April 2014. The registration number is DRKS00004462. For deviations from the protocol see *Appendix 6.4*.

6.5 COMPLIANCE WITH REPORTING STANDARDS

The methods, results and discussion collectively report all items required by the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) statement (Collins et al. 2015, Moons et al. 2015). The

[6] This appendix includes the German letter and an English translation of its content.
[7] Link: https://drks-neu.uniklinik-freiburg.de/drks_web/ [last accessed 13 May 2016]
[8] Link: http://apps.who.int/trialsearch [last accessed 13 May 2016]

completed (prediction model development) checklist for the study can be viewed in *Appendix 6.5.*

6.6 STUDY METHODS

The relevant general methodological concepts were introduced in *Chapter 2*. The design of the study was, as far as possible, based on the most recent methodological guidance and publications on prognosis research as outlined in *Chapter 2* (see *section 2.11*).

6.6.1 OVERWIEW OF STUDY DESIGN

This was a prospective, single-group cohort study (Moons et al. 2009; Royston et al. 2009). *Figure 6.1* illustrates the study design in terms of the key elements of prognosis research (as introduced in *Chapter 2 Figure 2.3).* The phase of prognostic model research was development.

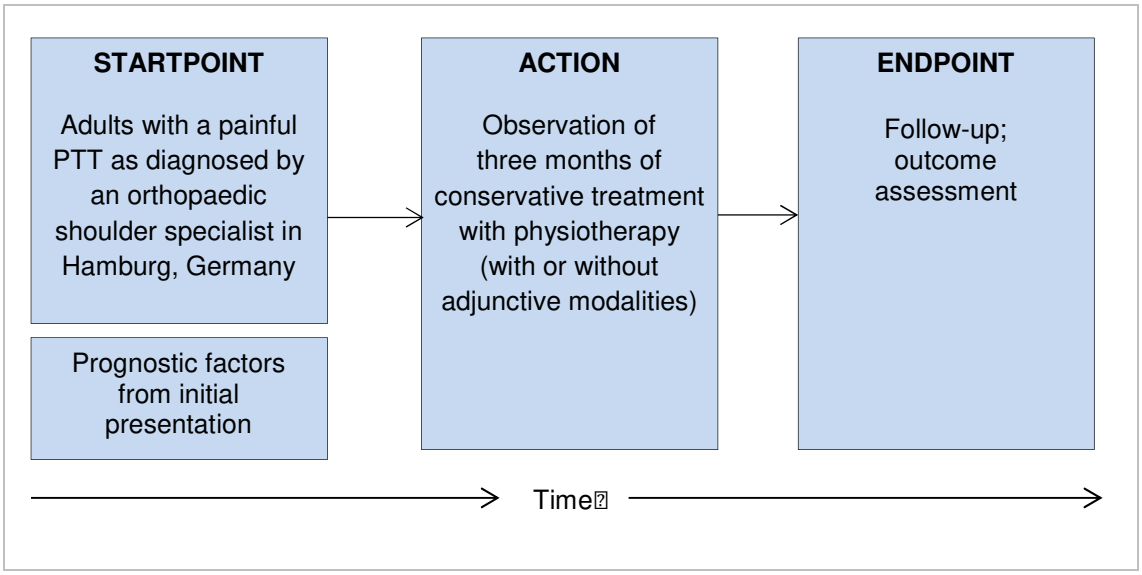


Figure 6.1: Outline of study design

6.6.2 GENERAL SETTING

The study was conducted in Hamburg, Germany.

6.6.3 STUDY PERSONNEL IN CONTEXT

The study involved collaboration with a shoulder surgeon and physiotherapists in clinics across the Hamburg area.

6.6.3.1 Shoulder surgeon

The shoulder surgeon, Dr Andreas Betthäuser, is a certified instructor in ultrasonographic shoulder diagnosis (Deutsche Gesellschaft für Ultraschall in der Medizin (German Society for Ultrasonography in Medicine), DEGUM). He conducts approximately 4,000 shoulder scans annually in his single-handed secondary care clinic (Schulter-Zentrum Hamburg; www.schulter-zentrum.com). Dr Betthäuser conducted all recruitment into the study, performed and interpreted all initial and follow-up assessments, including ultrasound scans, and, in keeping with his role as the "responsible treating doctor" (as required by German ethics regulations), monitored the treatment phase.

6.6.3.2 Physiotherapists

The physiotherapy took place in collaborating physiotherapy practices whose staff also undertook some patient-specific administrative tasks, namely collection of the signed informed consent forms and completion of physiotherapy report forms. The practices did not have to fulfill any specific requirements in terms of facilities, equipment or specialist training of staff, but their leads were expected to ensure that all staff involved would have sufficient experience and expertise in the physiotherapeutic treatment of patients with PTTs and impingement-related shoulder complaints to comply with the protocol, and to provide written evidence of their staffs' indemnity as required by Teesside University's indemnity policy.

As outlined in *Chapter 4* (*section 4.8*), eleven potentially eligible practices were initially selected based on my own knowledge, recommendations by Dr Betthäuser and other colleagues, and complementary web searches. I telephoned the practice leads to gauge their interest and evaluate their practices' eligibility. Nine were willing to collaborate (two lacked the time), and seven practices (comprising nine locations) were eligible. (The other two focused on osteopathic rather than physiotherapeutic care.) As previously indicated (*Chapter 4*), the leads of these seven practices confirmed that the protocol, as outlined, complied well with their usual approach to the treatment of patients with impingement-related shoulder pain. Staff in the seven practices were subsequently sent written information (in German, and therefore not appended to this

thesis, and materials including: blank copies of the consent and physiotherapy report forms, prepared reply envelopes, contact details for Dr Betthäuser and me, and a sample "study pack" (see further)). A face-to-face meeting was then arranged at each site (between May and July 2012) to provide opportunities for discussion. Following the respective meetings, the practice leads each signed a statement (see *Appendix 6.6*) confirming that they had been informed about the study and had agreed to collaborate, and that their staff were appropriately indemnified.

Patients were free to choose which physiotherapy practice they would attend. The list of the seven collaborating practices was provided, but for various reasons, often logistical, patients sometimes expressed a preference to be treated elsewhere. To accommodate this, and optimise recruitment, these additional practices (numbering 24) were contacted and incorporated *ad hoc*, but following the same processes detailed above[9].

Throughout the study I maintained regular contact with the practice leads and their clinical staff, mainly by telephone or email.

6.6.4 STUDY POPULATION

The population of interest was adults (18 years) with painful, ultrasonographically diagnosed PTTs presenting to Dr Betthäuser. The eligibility criteria were developed to restrict inclusion to patients whose shoulder pain could reasonably be linked to the presence of a PTT, while avoiding over-exclusivity. The eligibility criteria are presented in *Textbox 6.1*.

[9] The incorporation of further physiotherapy practices reflects a deviation from protocol (see *Appendix 6.4*).

Textbox 6.1: Study eligibility criteria

***Inclusion*:**

- Patients with (local) shoulder pain in the presence of an atraumatic (ultrasonographically detected) PTT
- Clinical signs of shoulder impingement (e.g. painful arc, positive impingement tests)
- Adults (□18 years)
- No restrictions on sex
- Agreement on conservative treatment
- Ability to speak and comprehend the German language
- Agreement to participate (signed informed consent)
- Anticipated availability for follow-up (living in area of Hamburg)
- Agreement to physiotherapy in a collaborating practice

Exclusion:

- Presence of an FTT at the affected shoulder
- Previous substantial shoulder trauma (e.g. shoulder dislocation, fractures)
- Previous surgery for the affected shoulder
- Previous surgery in the shoulder area that may be causal of or contributory to the current problem (e.g. surgery for breast cancer)
- Clinically relevant glenohumeral degeneration or disease (e.g. frozen shoulder)
- Current glenohumeral septic arthritis
- Clinically relevant acromioclavicular arthritis (e.g. local tenderness, positive provocation tests)
- Clinically relevant calcific tendinitis
- Ultrasonographic evidence of LHB tendon subluxation/ dislocation
- Referred pain from the cervical spine region
- Multisite musculoskeletal pain
- Systemic disorders, diseases or comorbidities as potential sources of (the current) shoulder pain (e.g. breast cancer, rheumatoid disease, inherited disorders (e.g. Marfan syndrome, Ehlers-Danlos syndrome)), or as impairing treatment (e.g. cancer, cardiac insufficiencies)
- Neurological disorders or deficits as potential sources of (the current) shoulder pain or impairing assessment and treatment (e.g. hemiplegic shoulder)
- Worker's compensation claims
- Unwillingness or inability to give informed consent (e.g. cognitive or intellectual impairments)

6.6.5 ASSESSING ELIGIBILITY

The assessments conducted were standard for Dr Betthäuser's practice.

6.6.5.1 Physical assessment

Dr Betthäuser conducted the physical assessment based on recommendations by the German Society for Shoulder and Elbow Surgery (Deutsche Vereinigung für Schulter- und Ellbogenchirurgie, DVSE) (DVSE 2012). In *Appendix 6.7*, an overview is provided of the recommended components and selection of tests. Dr Betthäuser chose the tests based on the individual patients' presentations. All tests were carried out on both shoulders. Further aspects of the assessments are provided with the study procedures.

6.6.5.2 Ultrasonographic assessment

Dr Betthäuser determined the presence of each PTT ultrasonographically using an ultrasound unit in appliance class 3 (the highest classification) according to DEGUM (http://www.degum.de/service/geraetelisten.html) and a linear transducer with a resolution of □10 MHz and a width of □40 mm.

The assessment was based on the standards by DEGUM and the German Society for Orthopaedics and Trauma (Deutsche Gesellschaft für Orthopädie und Unfallchirurgie, DGOU) (Konermann & Gruber 2007). As the minimum standard, the ventral, dorsal, superior and lateral shoulder regions were assessed both in the transveral and longitudinal plane. The assessment was tailored to the individual patient and could thus involve further planes. The assessment always included both a static and a dynamic evaluation.

Diagnosis of a PTT was based on the criteria by Hedtmann & Fett (1995, 2002). The key criteria for the presence of a rotator cuff defect were alterations of structure (echogeneity) and form (diameter, reversal of contour). In distinction to a PTT, an FTT was marked by the absence of the depiction of the rotator cuff (discontinuity of the rotator cuff). *Figure 6.2* shows three examples of rotator cuff ultrasound scans: an intact cuff, a supraspinatus PTT and a supraspinatus FTT. (All scans show a right shoulder. All three patients were male and between 59 and 62 years old).

□

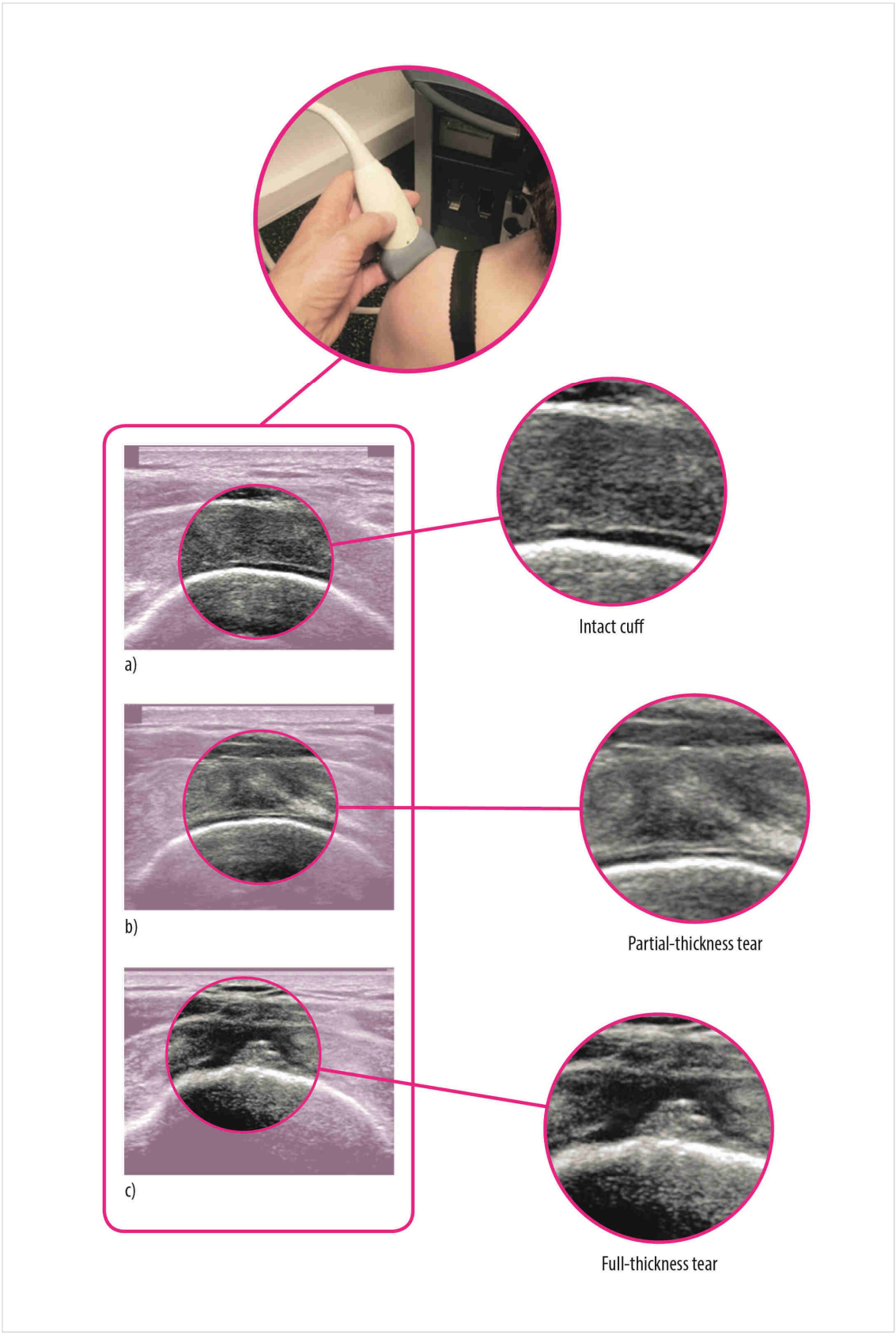


Figure 6.2: Example US scans of the rotator cuff
(US scans and picture provided by Dr Betthäuser)

6.6.6 TREATMENT

All study participants were followed over approximately three months of standard care with physiotherapy at one of the collaborating physiotherapy practices, with or without adjunctive treatment such as oral medication or local steroid injections as deemed clinically appropriate.

The physiotherapy treatment followed the broad protocol as described in *Chapter 4*. Beyond that, and in keeping with the ethos of an observational study, the type, content and amount of treatment were unregulated. Participants were advised on the amount and content of physiotherapy on an individual basis by their treating therapist and in agreement with Dr Betthäuser. All treatments were delivered in compliance with German national healthcare regulations (see *Chapter 2 section 2.10.4*). Details were documented.

Dr Betthäuser implemented, monitored and documented any adjunctive medical treatment and any modifications of the treatment regime that were considered necessary. Participants were free to consult Dr Betthäuser during the treatment phase if they felt in need of further advice or adjunctive medical treatment, which was in line with usual practice.

6.6.7 PROGNOSTIC FACTORS

The 10 candidate prognostic factors for this study (fully described and defined in *Chapter 5*) were: age, sex, physical demands, disability, pain, history of shoulder pain, symptom duration, diabetes, smoking and pain catastrophizing. All were assessed during the baseline assessment, some by patient-completed questionnaires that formed part of a paper-based initial questionnaires package (*Appendices 6.8, 5.5* (WORC) and *5.6* (PCS)), and others by Dr Betthäuser's standard clinical history interview.

6.6.8 ADDITIONAL BASELINE CHARACTERISTICS

Additional baseline characteristics, collected not for modelling purposes but to enhance characterisation of the study sample, were: "affected tendon" (supraspinatus, infraspinatus, supraspinatus *and* infraspinatus, or any other), "involvement of dominant arm" (yes or no) and "work status" (full-time, part-time, sick leave, retired, or not working (any other reason)). As with the candidate prognostic factors, some of these

data were derived from a patient-completed questionnaire (*Appendix 6.8*) and others from Dr Betthäuser's standard clinical history.

6.6.9 OUTCOME MEASURES AND MEASUREMENT

6.6.9.1 Primary outcome measure

The primary outcome measure was the change in the WORC score (WORC_change) between baseline (WORC_1) and follow-up (WORC_2). The WORC is described in *Chapter 5* (*section 5.5.2.1*), because it was also used as a prognostic factor. Participants completed the WORC_2 as part of a paper-based "follow-up questionnaires" package which included the WORC (*Appendix 5.5*) and the Global Perceived Change (GPC) scale (see further). Completion of the WORC_2 was standard at follow-up assessments in Dr. Betthäuser's practice, but for the purposes of the study, alternative provision was also made for postal completion, as required (see timing of outcome assessment, below).

6.6.9.2 Secondary outcome measures

Global Perceived Change (GPC)

The participant-perceived overall change of the shoulder problem was assessed by a GPC scale (Kamper et al. 2009) (see *Appendix 6.9*). GPC scales are widely used as an overall measure of patients' perceptions of how their health status has improved over a period of treatment or time. While GPC scales may vary in their design, they generally ask patients to rate the perceived change in their health status by comparing its current state to that at a defined, previous point in time. There are some criticisms of GPC scales, specifically that they may be prone to recall bias and that current status may influence ratings more than change (Kamper et al. 2010). However, their face validity is high, the limited evidence available on their measurement properties suggests that they are adequately reproducible and sensitive to change, and they are strongly correlated with various other PROMs such as the Roland Morris Disability and the Oswestry Low Back Pain Disability Questionnaire (Kamper et al. 2009). GPC scales are a recommended core outcome measure for use in chronic pain trials (Dworkin et al. 2005, Kamper et al. 2009) and are the most widely used patient-based anchor in studies determining the MID of outcome measures (Brozek et al. 2006, Jevsevar et al. 2015, Revicki et al. 2008, Turner et al. 2010, Wright et al. 2012).

Consequently, a GPC scale was also used in the MID analysis of the WORC (see *Chapter 7*).

Based on an investigation of different numbers of response options in rating scales by Preston & Colman (2000), Kamper et al. (2009) postulated that "scales with 7 to 11 points appear to offer the best compromise between patient preference, adequate discriminative ability, and test-retest reliability" (p. 168). The scale was therefore constructed around a central ("unchanged") point, with equal numbers of response points (-3 to +3) on both sides (Kamper et al. 2009 p. 168). In addition to the numbers, written descriptors were assigned to each response point. The phrasing of the question (i.e. the anchor statement) is critically important if the scale is to measure what it is intended to measure (Kamper et al. 2009). To this end, there was explicit mention of the health condition [shoulder problem], the construct of interest [overall change] and the reference time point for the assessment of change ["since your first assessment with Dr Betthäuser"] (Kamper et al. 2009).

As with the WORC_2, GPC rating was done as part of the follow-up assessment but could alternatively be done postally.

Tear progression

Tear progression was defined as progression from a PTT to an FTT (yes or no) and was measured ultrasonographically by Dr Betthäuser. Presence of an FTT was based on the ultrasonographic criteria described above (*section 6.6.5.2*).

Tear progression was primarily assessed as a means of determining the safety of a three-month phase of conservative treatment with physiotherapy in patients with PTTs. I had initially intended to measure progression on a continuous scale (in mm or % change in depth), but a preliminary review of the literature cast doubt on the validity and reliability of ultrasonographic cuff tear measurement in terms of either tear depth or area (Bryant et al. 2002, Iannotti et al. 2005, Kim et al. 2011, Teefey et al. 2004).

Adverse events

The treating physiotherapists were asked to document any adverse events that they considered as related to the physiotherapy treatment. This was done by an open question on the physiotherapy report form (*Appendix 4.2*): "Did your patient report any problems such as exacerbations of symptoms or side effects?" and a blank space. The question was deliberately kept broad to allow for any experienced adverse events.

6.6.9.3 Blinding of outcome assessment

In strict terms, assessment of the primary outcome (WORC_change) could not be blinded to knowledge about the prognostic factors, because both the outcome and all prognostic factors were patient-reported. In practice however, patients were not aware which of the many parameters evinced at baseline were candidate prognostic factors. The forms were completed in a separate room, i.e. not in the presence of Dr Betthäuser.

6.6.9.4 Timing of outcome assessment

Follow-up assessments were conducted as close as possible to three months after the baseline assessment. This complies with Dr Betthäuser's usual practice. Some flexibility was allowed, but four months was considered the longest acceptable interval. For patients who would be unable to attend within this timeframe, follow-up questionnaires were sent by post instead, aiming for return (in a postage-paid envelope) at around three months after the baseline assessment.

6.6.10 PATIENT INFORMATION AND INFORMED CONSENT

All potential participants were verbally informed about the study by Dr Betthäuser. Those who expressed potential interest were given a fuller explanation, supplemented by a written patient information sheet (PIS, *Appendix 6.10*) and an informed consent form (*Appendix 6.11*). The sheet and form were developed following guidance from various sources (e.g. Teesside University 2011, Health Research Authority 2011) and the requirements of the ethics commission of the Hamburg Medical Council.

In order to give sufficient time to consider participation, all eligible participants were asked to take the form with them and sign it, if they still wished to do so, during their first physiotherapy session.

6.6.11 POTENTIAL RISKS TO THE INTERESTS OF PARTICIPANTS

The risks to the interests of participants were considered minimal as there was no deviation from standard clinical practice. As part of his usual practice, Dr Betthäuser informed all participants of any potential risks or harms related to any treatment he initiated or provided. Regarding the physiotherapy treatment of patients with PTTs, there are no known serious risks or harms. Participants were routinely informed that temporary increases in symptoms could occur as a reaction to the physiotherapy

treatment, and that in case of any such reaction, the treating therapist would be the primary person to manage this. Participants were further informed that they could at any time contact Dr Betthäuser, who, as routine, monitored the whole treatment phase. Dr Betthäuser also initiated, monitored and documented provision of any adjunctive medical treatment (e.g. oral pain medication, local steroid injections to the shoulder), or any modifications of the treatment regime if considered necessary.

6.6.12 STUDY PROCEDURES

In the following section, the study procedures are described. *Figure 6.3* provides an outline of the procedures from the baseline (initial) assessment (the study's startpoint) to the follow-up assessment (the study's endpoint).

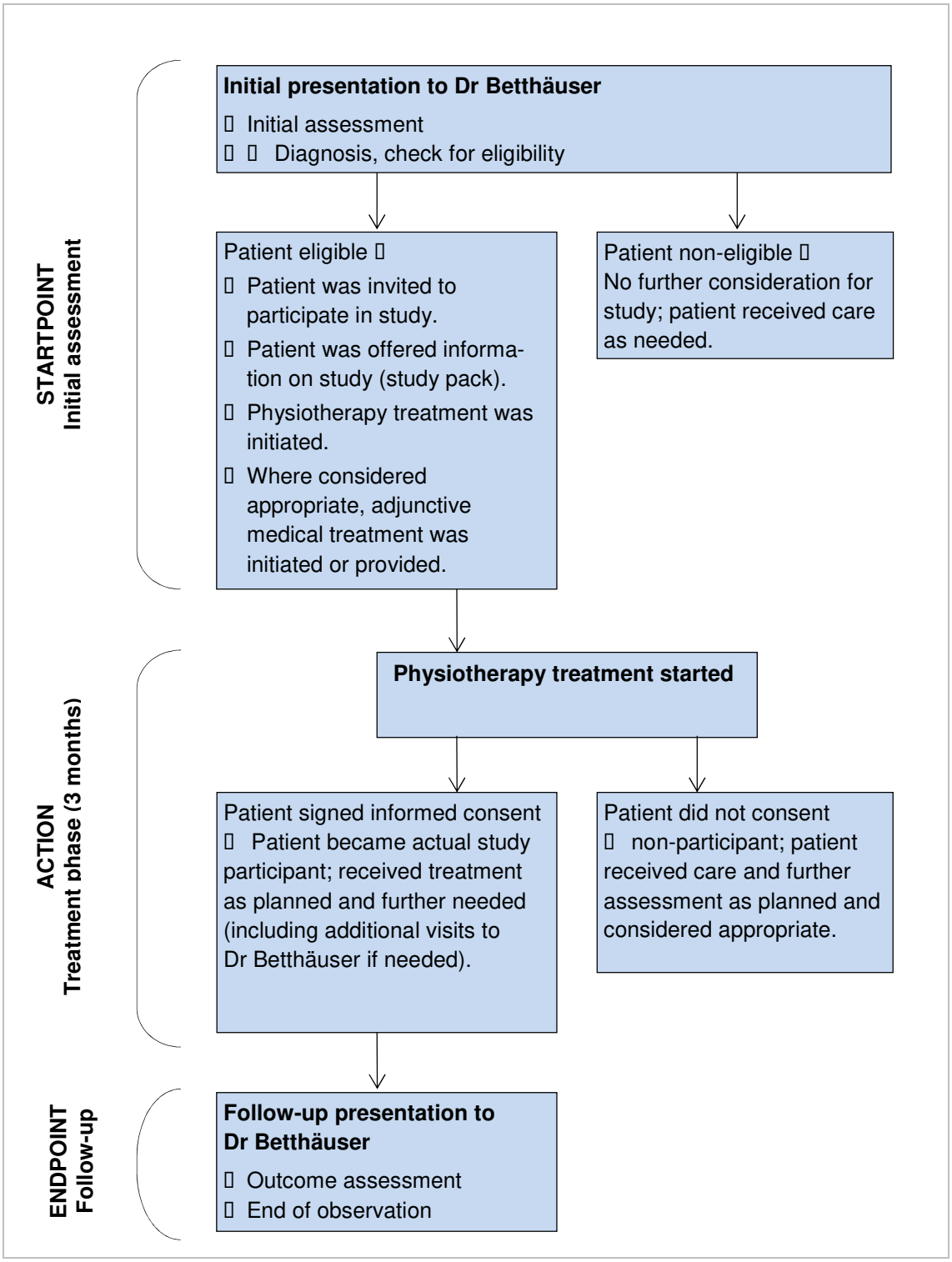


Figure 6.3: Overview of study procedures

6.6.12.1 Baseline assessment and diagnosis (startpoint)

At baseline, Dr Betthäuser assessed and diagnosed all patients (as described in *section 6.6.5*). The assessment included:

- a subjective assessment of the patient's history and current complaints, comprising a one-to-one interview and the completion of participant-completed questionnaires (see *sections 6.6.7* and *6.6.8*). Evaluation of the candidate prognostic factors was part of this assessment.
- a physical assessment
- an ultrasonographic assessment

6.6.12.2 Recruitment

Throughout the recruitment period, consecutive patients were screened for eligibility. Eligible patients were informed about the study and invited to consider participating. Those who were interested were given a study pack in an envelope labelled with a unique study ID. The study pack included:

- a PIS labelled with a unique study ID (*Appendix 6.10*)
- an informed consent form without the study ID (*Appendix 6.11*)
- a list with the names and contact details of the primary collaborating physiotherapy practices (not included in appendices for reasons of confidentiality)
- a physiotherapy report form labelled with the study ID (*Appendix 4.2*)

Dr Betthäuser documented the unique study ID and asked the patient to arrange a follow-up assessment at three months after the initial presentation (as is his standard practice). He also initiated or provided further medical treatment (e.g. oral medication, corticosteroid injections), if appropriate, and provided the patient with a prescription for physiotherapy.

Thus, all eligible patients who were interested in the study left Dr Betthäuser's practice with a study pack and a physiotherapy prescription. They were invited to read the information and were informed that they could contact Dr Betthäuser or me in case of any further questions about the study. They were asked to keep the study pack together with their prescription, and to take it with them when they saw their physiotherapist.

Patients wishing to participate were asked to choose one of the collaborating physiotherapy practices for their treatment. If they indicated that they would prefer a

different practice, they were asked to do the same but where additionally asked to either contact me or ask the treating therapist to contact me for the relevant information about the study and to implement collaborator status (see *section 6.6.3.2*).

6.6.12.3 Treatment

During the first physiotherapy treatment session, the treating physiotherapists asked the patient if he or she had decided to participate in the study. If so, the consent form was completed and returned, by the physiotherapist, to Dr Betthäuser, who informed me. If a patient had forgotten to bring the study pack, but wished to participate, the physiotherapist provided a spare blank form, and contacted me for the study ID. The physiotherapist added a note to the patient documentation when a patient had given consent, and attached the report form containing the study ID to the patient documentation, ready for completion at the end of the study observation period.

Upon receipt of the signed consent form, I confirmed eligibility according to the data documented by Dr Betthäuser. Any uncertainties were resolved through discussion with Dr Betthäuser. Receipt of a signed consent form and confirmation of eligibility by me marked a patient's entry into the study. Participants then received treatment as described (see *section 6.6.6* and *Chapter 4 section 4.6*). On completion of their original physiotherapy prescription, participants were individually advised on the number and type of follow-up prescriptions by the treating physiotherapist in agreement with Dr Betthäuser.

6.6.12.4 Follow-up assessment (endpoint)

When the physiotherapy treatment phase ended, the physiotherapists reminded their patients of the follow-up assessment with Dr Betthäuser, completed the physiotherapy report forms and sent them to Dr Betthäuser. The follow-up assessment by Dr Betthäuser included:

- a subjective follow-up assessment of the patient's complaints comprising the completion of the WORC_2 and the GPC scale (see *sections 6.6.9.1* and *6.6.9.2*)
- a clinical re-examination and
- an ultrasonographic re-examination

The follow-up assessment marked the end of the observation and involvement of the patient in the study. Follow-up was considered completed if as a minimum the primary outcome measure, i.e. the WORC_2, was completed.

6.6.13 DATA COLLECTION AND MANAGEMENT

6.6.13.1 Questionnaire data

Participants were asked to carefully read the instructions provided in the questionnaires, to complete them in full, and to ask for clarification if they had any questions. Standard ballpoint pens were provided for the completion of the questionnaires in Dr Betthäuser's practice. I printed all questionnaire paper copies myself to ensure that the length of all VAS scales (WORC and pain) were exactly 100 mm long, as printing or copying of VAS scales can lead to a distortion of the length of the scales (Schomacher 2008).

6.6.13.2 Evaluation of questionnaire data

Specific attention was given to the evaluation of the data that were measured by VAS, i.e. WORC and baseline pain. All measurements were done with the same standard ruler. The length of all VAS scales was double-checked. The distance between the left end of the line and the participant's mark on the line was measured to establish the score. Measurement was in mm, and to the nearest mm, or, where this was unclear, to the higher mm. All measurements were done twice to minimise the risk of error. When a rating was difficult to read, it was independently assessed by a second person (either Dr Betthäuser or Dr Hanchard) to reach consensus where possible. Otherwise, the rating was considered missing.

6.6.13.3 Interview data

As described in the sections on the prognostic factors, additional baseline data and outcome measures (*sections 6.6.7 to 6.6.9*), some of the data were collected verbally as part of Dr Betthäuser's standard interviews. Dr Betthäuser documented all his assessments in his patient database (see further).

6.6.13.4 Data storage

All paper-based patient documents were stored securely in Dr Betthäuser's practice, in view of his responsibility as the treating doctor. Study IDs were removed from postal questionnaires and physiotherapy report forms, thus removing any link between named patient data and study data. The standard documentation maintained by the

collaborating physiotherapy practices was outside the scope of governance for this study.

6.6.14 SAMPLE SIZE CONSIDERATIONS

Due to the multivariable nature of prognostic development studies, it is difficult to estimate the required sample size (Moons et al. 2009). Indeed, there are no formal methods (based on either power calculations or adequate precision of estimation of effects) to determine the effective sample size. A commonly used rule of thumb is that there should be at least 10 outcome events (for binary or time to event outcomes) or individuals (for continuous outcomes) per prognostic factor (Peduzzi et al. 1996), with the number of outcome events relating to the smallest group when the outcome is categorical (Bouwmeester et al. 2012). However, more recently it has been proposed that this guideline might be too conservative and fewer than 10 events per prognostic factor might provide adequate accuracy and precision of estimation of effects. In a comprehensive and rigorous simulation study Vittinghoff and McCulloch (Vittinghoff & McCulloch 2007) reported that CI coverage, Type 1 error rate, and relative bias were comparable in scenarios with 5-9 events per prognostic factor versus those with 10-16 events per factor. Following this work, and in order to analyse the WORC on a continuous scale[10], I based the minimum sample size of my study on a requirement for 5-9 individuals per prognostic factor. With 10 prognostic factors, which was considered achievable, this required (5 to 9)*10 = 50 to 90 participants which, inflated by 20% to allow for losses to follow up, resulted in a target of 60 to 108 participants.

6.6.15 MISSING DATA

Missing data were documented for all prognostic factors and outcomes. Missing values for the two multi-item PROM questionnaires, the WORC and the PCS, were replaced. As no standard missing rules were available for the WORC from the literature, I decided to replace missing values by the mean of the domain to which the missing value belonged. In case of missing data for the PCS, I followed the approach suggested by the primary originator of the PCS, Prof Peter Sullivan, whom I contacted in June 2014. He responded: “If there are one or two items missing, we compute the mean of items that were completed and substitute the mean for the missing values. The internal consistency of the measure is sufficiently high to make this approach acceptable.” (Personal communication 02/06/2014). I did not replace missing

[10] The analysis of the WORC on a continuous scale reflects a deviation from protocol; see Appendix 6.4

data for the WORC and PCS in cases where the complete questionnaire data were missing.

No further missing values were replaced. The decision about which method to apply for dealing with any further missing data, i.e. whether to conduct a complete-case analysis or to impute missing data, was made after the inspection of the data, but prior to conducting the analysis. I intended to consider in particular the amount of, but also potential reasons for, missing values. The chosen approach to the model selection (AIC approach, see further) required identical samples (see further). The analysis was eventually conducted on a complete-case basis, which is further addressed with the results.

6.6.16 DATA CHECKS

The study data were kept on a Microsoft Office 2011 (for Mac) Excel database. Prior to the statistical analyses, data were scrutinised for completeness and possible data abstraction errors. Comprehensive double checks were conducted to ensure that all data had been abstracted correctly. Any identified mistakes were corrected.

6.6.17 STATISTICAL ANALYSIS

6.6.17.1 Handling of candidate prognostic factors in the analyses

The process of selecting candidate prognostic factors for inclusion in the study is described in detail in *Chapter 5*. All initially considered factors were included in the multivariable analysis, i.e. there was no further selection prior to the multivariable modelling. All continuous factors were analysed on their continuous scale, i.e. none were categorised for the multivariable modelling. The PCS and WORC scores were treated as continuous measurements. All non-continuous factors were binary variables. Their numerical coding is presented with the results. All continuous factors were modelled as linear (see further for checks of assumptions).

6.6.17.2 Handling of the outcome in the analyses

As previously described, the outcome (WORC_change) was also analysed on a continuous scale. The data was analysed as linear (see further for checks of assumptions).

□

6.6.17.3 Statistical analysis methods

The primary approach to the multivariable modelling of the candidate prognostic factors (as independent variables) in relation to the outcome (i.e. WORC_change as dependent variable) was a linear regression analysis (Burnham & Anderson 2002, Miles & Shevlin 2001, Royston et al. 2009).

6.6.17.4 Checking assumptions of linear regression

Satisfaction of the assumptions of linear regression, i.e. normality, homoscedasticity, linearity and independence was checked for each model by inspecting the scatterplot of the standardized residuals against the standardized predicted values (residual plot) (Miles & Shevlin 2001 pp. 84–112).

6.6.17.5 Approach to model selection: Akaike's Information Criterion (AIC) approach

The general aim of multivariable modelling analysis is to find a parsimonious model, i.e. "one that explains the most variance in the dependent variable containing the fewest number of independent variables." (Miles & Shevlin 2001 p. 38). This can be achieved through a variety of techniques, many of which, though, such as forward, backward and stepwise regression rely entirely on statistical criteria with arbitrary cut-offs for "significance" (p values) (Miles & Shevlin 2001 p. 38). As stated in the prognostic systematic review report (*Chapter 3*), these automated, "p-value-based" approaches, but in particular stepwise regression, have consistently been linked to biased predictions (Flom & Cassell 2007, Harrell 2001 pp. 56–8, Miles & Shevlin 2001 pp. 38–9, Sainani 2013). As Miles & Shevlin (2001 p. 38) state, "The basic problem [of stepwise regression techniques] is that we are asking the computer to make a decision regarding which variables are important, when the computer has no idea about the theory that may determine which variables are important." (p. 38). I avoided these issues by using Akaike's Information Criterion (AIC), an information-theoretic approach which was developed by Hirotogu Akaike in the 1970s (Anderson 2008 p. 55, Burnham et al. 2011).

Information-theoretic approaches to model selection differ from the automated selection methods such as forward, backward or stepwise regression in important ways. One is that selection is based on the comparison of multiple (candidate) models representing alternative hypotheses rather than on a single global cluster of prognostic factors (Burnham 2004, Burnham et al. 2011). A second principle is that the selection

is based on an information-theoretic criterion (such as the AIC, see further), which provides “numerical values that represent the scientific evidence” for each model (Anderson 2008 p. 64), but which does not provide any test statistics such as p values and does not involve any arbitrary decisions based on statistical significance (Anderson 2008 p. 64).

The AIC approach is based on Kullback-Leibler (KL) information (Burnham 2004, Burnham et al. 2011). Reflecting the perspective that models never reflect “full reality”, i.e. that they are approximations (Anderson 2008 p. 27), KL information “represents the *information lost* when model g_i [a model] is used to approximate full reality.” (Burnham et al. 2011 p. 24) or, equivalently, the “distance” between the approximating model and full reality (Anderson 2008 p. 53). *Figure 6.4* (see left side) illustrates KL information (d_X), with f representing full reality and g_X representing a set of models (Anderson 2008 p. 24, Burnham et al. 2011 fig. 1). The full reality is inherently unknown and cannot be estimated, but is the same for all models, and can thus be removed (Burnham et al. 2011). The AIC approach uses a criterion (the AIC) to identify the model within a set of candidate models that is “closest to full reality” (Burnham et al. 2011 p. 24). This is then termed the “best” model and used as the reference model to which the other candidate models are compared. The AIC value provides an estimate of the KL information, with smaller AIC values representing a smaller loss of information, i.e. a closer approximation to full reality. Thus, the model with the smallest AIC value (AIC_{MIN}) is considered the one that is „closest to full reality“ (Burnham et al. 2011 p. 24).

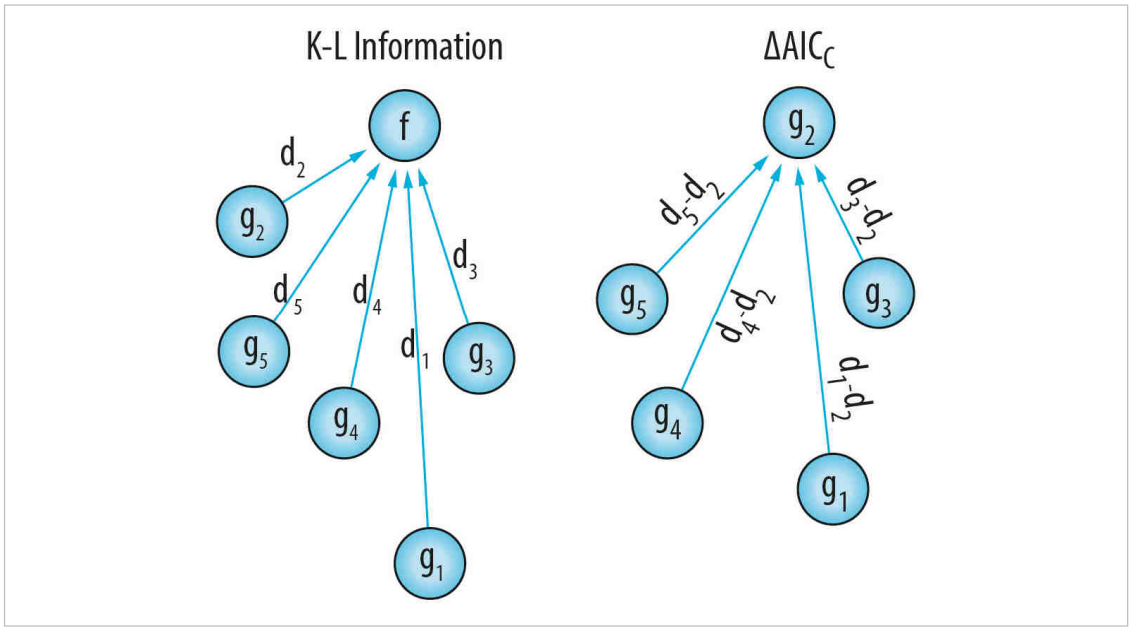


Figure 6.4: KL information and AIC differences
(modified from Burnham et al. 2011 fig. 1)

The candidate models should be defined a priori; they represent different “working hypotheses” and should ideally be derived by sound theoretical reasoning, i.e. by consideration of “the science of the matter, experience and expertise” (Burnham & Anderson 2002 p. 96). Their number may vary, though it is recommended that it should usually be limited to a few, and should, as a general rule, not exceed the size of the study sample (Anderson 2008 p.62). As AIC values are functions of sample size, the models must be based on identical datasets to allow for their unbiased comparison (Anderson 2008 p. 63).

Comparison of the other candidate models to the best model is done by calculating their distance to the best model, i.e. by calculating AIC differences ($\Delta AIC = AIC - AIC_{MIN}$), which allows for a ranking of the models according to the amount of their distance to the best model. Smaller ΔAIC values represent a closer approximation to the best model, thus providing an indication of a better relative fit. Thus, whereas the AIC provides an estimate of the distance between each model and full reality, the ΔAIC is the difference between each model and the best (Anderson 2008 p. 84). *Figure 6.4* (see right side) illustrates the basic approach to AIC model selection (Burnham et al. 2011 p. 25).

AIC accounts for the number of prognostic factors by “penalising” models with larger numbers of prognostic factors, thereby favouring parsimony (Anderson 2008 pp. 58–9). A version of AIC for small samples (second-order bias-correction), termed AIC_C, is available and strongly recommended for use in cases where the sample size (n) in relation to the number of prognostic factors (K) is small (suggested ratio $n/K < 40$) (Anderson 2008 pp. 60–1). AIC_C was thus used for the analysis of the models in this study.

It is important to consider that AIC_C (and AIC) values are relative rather than absolute, and “on the scale of information” (Anderson 2008 p. 84). This means that the values themselves are uninterpretable (Burnham 2004). Rather, their use is limited to comparing models within a defined set of models (Burnham & Anderson 2002 p. 71). Moreover, though AIC will always select a best model among a set of models, this does not mean that it is necessarily a good model (Burnham & Anderson 2002 p. 62). Consequently, it has been suggested that the worth of the best or the global (full) model be assessed by e.g. a goodness-of-fit test, analysis of residuals or the adjusted R^2 (Anderson 2008 pp. 67–8).

For the interpretation of the ΔAIC of any model in relation to the best model, Burnham et al. (2011) have proposed considering models with ΔAIC values < 4-7 as “plausible” alternatives to the best model, models with ΔAIC values of 9-11 as having “relatively

little support", models with □AIC values > about 14 as "implausible" alternatives to the best model (p. 25), and models with □AIC values > 20 as having "essentially no empirical support" (p. 25).

Further methods have been proposed for the scaling and comparison of each model within a set of candidate models (see e.g. Burnham & Anderson 2002 pp. 74–9). The analysis of the models in my study, though, was limited to the identification of the best model and the ranking of the other models in relation to it.

6.6.17.6 Candidate models

Prior to the analysis, I developed a set of candidate models. The choice of the models was based on clinical and theoretical considerations, which I discussed with Dr Betthäuser and Dr Hanchard. The intention was not to assess any possible combination of factors, but to select a limited number with a justification for why each might be of interest. In *Table 6.1*, the models are presented together with a brief rationale for each. The first model included all ten candidate prognostic factors (thus representing the "full" model), whereas the further models, all nested models (i.e. all factors were part of the full model), included between two and eight factors. As outlined in the table, selection considered various aspects such as the "potential for modification" or "effort of assessment". Thus, while I aimed to find a parsimonious model, I was also interested in considering aspects such as the effort (time) required to assess the prognostic factors, which would be highly relevant to clinical practitioners.

Table 6.1: Candidate prognostic models

No	Candidate model	N	Rationale
1	Age + sex + physical demands + disability (WORC) + pain + history of shoulder pain + symptom duration + Diabetes + smoking + pain catastrophizing (PCS)	10	Full model (all factors)
2	Diabetes + smoking + pain catastrophizing (PCS)	3	Potential for modification (could be modified (addressed) by some action (e.g. treatment)
3	Age + sex	2	Factors that cannot be modified
4	Age + sex + physical demands + pain + history of shoulder pain + symptom duration + diabetes + smoking	8	Type of assessment: “no questionnaires”
5	Disability (WORC) + pain catastrophizing (PCS)	2	Type of assessment: “questionnaires”
6	Diabetes + smoking	2	Type of factor: “bio(logical) factors”
7	History of shoulder pain + symptom duration	2	Background (patient history)
8	Pain + history of shoulder pain + symptom duration	3	Further models: pain-related factors (excluding pain catastrophizing)
9	Pain + pain catastrophizing (PCS)	2	Further models: pain and attitude towards pain

6.6.17.7 Model selection

I intended to compare and rank the models by the AIC_C approach as described above. I intended to select the best model identified, but to select more than one “final” model if there were plausible alternatives to the best model.

6.6.17.8 Adjusting for regression to the mean

Regression to the mean (RTM) describes the statistical phenomenon of a “tendency for subjects who score below average on a test to do better next time, and for those who score above average to do worse” (Hopkins 2002, no pagination). This was potentially relevant because the WORC was used both as a prognostic factor (WORC_1, measured at baseline) and as the primary outcome (change_WORC). Thus, the WORC_1 values were adjusted for RTM ($WORC\_1_{ADJ}$) prior to the prognostic modelling analysis, and the WORC_change scores were then corrected accordingly

(WORC_change$_{ADJ}$). The adjustment was done using the following equation: $x_{ADJ} = \bar{x} + p\,(x - \bar{x})$, where $\bar{x}$ denotes the mean of the cohort, p the pre-post correlation for the cohort, and x the individual participant's pre-test value (Linden 2013, "Controlling for RTM through data analysis" para. 2).

6.6.17.9 Model validation and further analyses

I did not pre-specify any method of validation or any further evaluations of model performance (e.g. calibration or discrimination) or other analyses (e.g. accounting for complexities in the data), but planned to decide on them based on the results. I intended to compare the SEE of the best model(s) with the MID estimate of the WORC derived from the sample data, and to internally validate any model with an SEE that is substantially lower than the MID. The determination of the MID estimate is presented separately in *Chapter 7*. Bootstrapping is recommended as the preferable approach for the internal validation of multivariable prognostic models (Steyerberg & Harrell 2016).

I intended to conduct the following exploratory subgroup analyses to account for the anticipated variability of some aspects related to the treatment: amount of physiotherapy treatment (number of treatment sessions), provision of injection (yes or no) and length of follow-up.

6.6.17.10 Period of analysis and software

The statistical analyses were conducted between June and November 2015 using IBM SPSS software (version 22). Both the "Regression" and "Generalized Linear Models" analysis options were used to obtain all relevant statistics including AIC_C values.

6.6.18 Presentation of statistical parameters

Baseline, candidate prognostic factor and outcome data, where available, were presented for all participants who completed the study. As a standard, these included means, standard deviations and ranges for continuous data; and numbers and proportions for binary data.

I intended to present the following statistical parameters:

- Model summary (for each model): AIC_C, standard error of the estimate (SEE) and adjusted coefficient of (multiple) determination (R^2_{ADJ}). The SEE was the primary parameter for the assessment of the models' precision. The R^2_{ADJ} was reported for

the purpose of providing a complementary measure of the models' performance only, but not for ranking the models.

- Model coefficient statistics (for each model): the regression constant (Constant) and the unstandardized regression coefficients (B) of all prognostic factors with their (95%) CIs.
- Model comparison: AIC_C, ΔAIC_C and SEE values

Because this research was in the developmental phase, the models were not intended to be used in practice, and therefore full equations were not reported.

All relevant data as well as the results of the statistical analyses were presented in text and tables.

6.7 RESULTS

6.7.1 Recruitment, flow of participants and sample size

Recruitment took place over the 22 months from December 2012 September 2014 inclusive. The last follow-up was completed in January 2015. *Figure 6.5* outlines the flow of participants through the study.

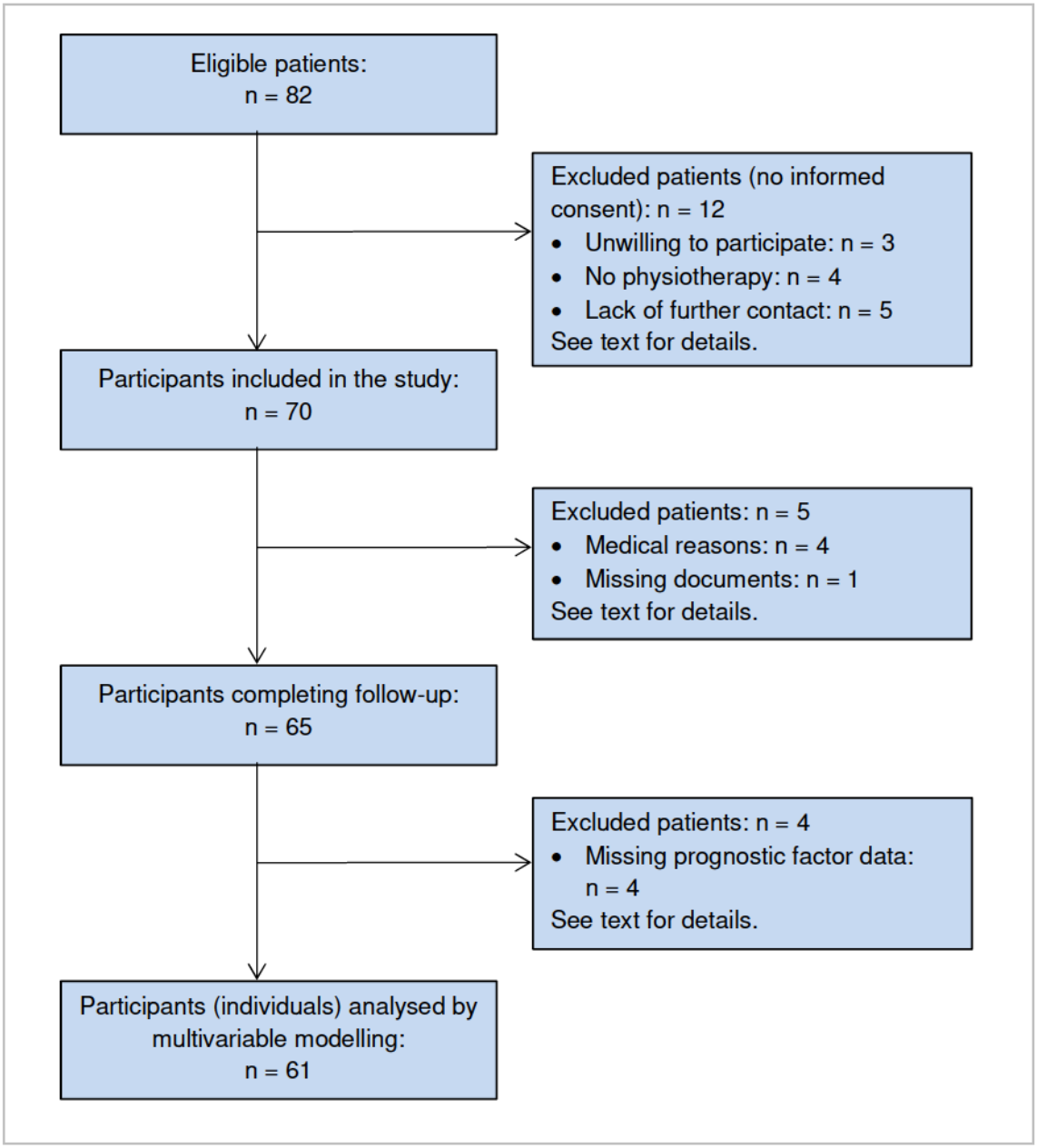


Figure 6.5: Flow of participants

By the end of September 2014, 82 patients had been invited to take part. For 12 of these, no consent form was received: three did not want to participate; four did not undergo any physiotherapy; five made no further contact with Dr Betthäuser or me. Thus, at that stage 70 participants had been recruited. Based on the rate of recruitment, it seemed that recruiting for another year (with the option of possibly including another 30-35 participants) would have no substantial influence on the precision of the prognostic models. In agreement with my supervisory team, it was thus decided to stop recruitment.

Of the 70 participants (70 shoulders) in the cohort, 65 (93%) completed the follow-up and 5 (7%) did not. In three cases this was due to newly arising medical reasons: one

participant injured himself in an accident shortly after inclusion in the study; one developed symptoms of a frozen shoulder; and another developed neurological symptoms. One further participant was retrospectively excluded because of a previous operation on the affected shoulder which was initially not documented. The follow-up questionnaires of one further participant were missing for unknown reasons and could not be obtained, necessitating exclusion.

Of the 65 participants who completed the follow-up, four were excluded from the analysis due to missing prognostic factor data. Thus, 61 (i.e. 87% of the enrolled participants and 94% of those who completed the study) were analysed by the multivariable modelling. This is further explained in the context of how I dealt with missing data (*section 6.7.1.5*). There were no complexities (e.g. unit of analysis issues) in the data.

6.7.2 Missing data

A small number of data items were missing for both some candidate prognostic factors and the outcomes for some of the 65 participants who completed the study.

6.7.2.1 Missing prognostic factor data

The missing data are presented separately for those factors that were assessed by a single-item measure (i.e. sex, age, duration of symptoms, pain, physical demands, history of shoulder pain, Diabetes and smoking) and those that were assessed by a multi-item measure (disability (WORC_1) and pain catastrophizing (PCS)).

Table 6.2 shows the amount of missing values for the eight candidate prognostic factors that were assessed by single-item measures. The six missing values related to two (3%) of the 65 participants; one of them had five missing values, the other one had one.

Table 6.2: Missing prognostic factor data: factors assessed by single-item measures
(n = 65)

Factor	Missing values	
	N	%
Age, yr	0	0
Sex, female/male	0	0
Physical demands, yes/no	1	2
Duration of symptoms, wk	2	3
Pain, mm VAS	1	2
History of shoulder pain, yes/no	1	2
Diabetes, yes/no	0	0
Smoking, yes/no	1	2

Table 6.3 shows the amount and key characteristics of missing values for the two candidate prognostic factors that were assessed by multiple-item measures, i.e. of missing values *within* the WORC_1 and PCS questionnaires. As the PCS was missing completely for three participants, the details in the table relate to the 62 participants for whom some values within the PCS were missing. The six missing values within WORC_1 were distributed as follows: one participant had two missing values, and four had one.

Table 6.3: Missing prognostic factor data: factors assessed by multiple-item measures

Factor (n)	Overall missing values, n (%) of all values*†	Items with missing values, list	Participants with missing values, n (%)
Disability (WORC_1)* (65)	6 (0.4)	3, 9, 13, 14, 16, 21	5 (8)
Pain catastrophizing (PCS)† (62)	1 (0.1)	13	1 (2)

**WORC: WORC: 21 items 65 x 21 values = 1365; †PCS: 13 items 62 x 13 values = 806 values*

6.7.2.2 Missing outcome data

Table 6.4 shows the amount and key characteristics of missing outcome data. The 11 missing WORC_2 values were distributed as follows: one participant had three missing values, one had two, and six had one. For 14 patients, who did not have a documented ultrasound scan within the follow-up period of the study, there were no tear progression data.

Table 6.4: Missing outcome data

Outcome (n = 65)	**Overall missing values, n (%) of all values***	**Items with missing values, list**	**Participants with missing values, n (%)**
Disability (WORC_2)*	11 (1)	4, 7, 8, 9, 13, 16, 21	8 (12)
GPC	1 (2)	n/a	1 (2)
Tear progression	13 (20)	n/a	13 (20)

**WORC: 21 items 65 x 21 values = 1,365; n/a = not applicable*

6.7.2.3 Approach to dealing with the missing data

As shown in *Tables 6.2* to *6.4*, the amount of missing data was small. For 61 (94%) of the participants, values were available for all prognostic factors. Regarding the primary outcome, 57 (88%) of the participants had no missing values for the WORC_2. I replaced all missing values for the WORC and PCS by applying the pre-specified rules (see *section 6.6.15*).

Following the replacement of the missing WORC and PCS values, four participants (6%) still had missing data for one or more of the prognostic factors. I decided not to replace any further missing values but to exclude these four cases from the analysis and to conduct the multivariable modelling analysis on a complete-case basis. As previously stated (see *section 6.6.17.5*), the AIC approach requires candidate models to be based on identical datasets. Excluding the four cases appeared legitimate as the impact on precision would be minimal. The missing data were too few to permit testing for differences between patients with and without them. However, since five out of the six missing values were only missing once, the reasons for missingness did not seem systematic. Thus, I considered the data from the complete cases as representative of the complete sample.

6.7.3 Baseline participant characteristics including prognostic factors

All of the participants had presented to Dr Betthäuser because of shoulder pain; all had been diagnosed with a painful PTT; and all fulfilled the predefined study eligibility criteria. The summary participant characteristics are presented in *Table 6.5*. These include the prognostic factors, all of which were assessed at baseline. Details on the prognostic factors, their definitions and measurements are provided in *Chapter 5*.

Table 6.5: ***Baseline participant characteristics - prognostic factors and additional characteristics***
Prognostic factors are listed by type of measurement (continuous/ categorical).

Characteristic (n)	Measurement	Values		
Continuous prognostic factors		x̄	SD	Range
Age (65)	yr	50	12	24-76
Disability (65)*	WORC_1 score	897	380	130-1660
Pain (64)	mm VAS	63	26	7-100
Symptom duration (63)	wk	36	49	1-250
Pain catastrophizing (62)*†	PCS score	15	9	1-37
Categorical prognostic factors		N	%	
Sex (65)	female	25	38	
	male	40	62	
Physical demands (64)	yes	41	64	
	no	23	36	
History of shoulder pain (64)	yes	35	55	
	no	29	45	
Diabetes (65)	yes	4	6	
	no	61	94	
Smoking (64)	yes	10	16	
	no	54	84	
Additional characteristics		N	%	
Affected tendon (65)	1. supraspinatus	63	97	
	2. infraspinatus	1	2	
	3. supraspinatus + infraspinatus	1	2	
	4. any other	0	0	
Dominant arm affected (65)	yes	46	71	
	no	19	29	
Work status (64)	1. full-time	41	64	
	2. part-time	11	17	
	3. sick leave	0	0	
	4. retired	10	16	
	5. not working (other reason)	2	3	

**Includes replaced values for missing data (see section 6.6.15); †PCS data were completely missing for three cases*

6.7.4 Treatment characteristics

6.7.4.1 Physiotherapy

Table 6.6 presents the key summary data related to the amount and duration of the physiotherapy treatment sessions. Only three participants (5%) had fewer than six sessions (one, two and three sessions, respectively), and 62 (95%) had six or more sessions.

Table 6.6: Physiotherapy – amount and duration

Aspect (n = 65)	**x̄**	**SD**	**Range**
Number of sessions	12	6	1-25
Duration of sessions (min)	28	13	20-80

Table 6.7 presents the summary data of the key components of the physiotherapy treatment, as documented by the treating physiotherapist in the physiotherapy report for each participant (see *Appendix 4.2* for the report form). For reasons of clarity, the domains are grouped into several categories such as “types of exercises”, “types of manual techniques” or “supplementary modalities”.

Table 6.7: Breakdown of physiotherapy interventions
Interventions are listed by general category and specific domain; domains are in descending order of use.

Category	**Domain** (n = 65)	**N**	**%**
Types of exercises	Strengthening exercises focused at rotator cuff muscles	52	80
	Scapula positioning exercises	47	72
	Stabilisation exercises	41	63
	Stretching techniques or exercises (shoulder/shoulder girdle)	36	55
	Strengthening exercises focused at shoulder girdle muscles	34	52
	Humeral head 'positioning' exercises	33	51
	Coordination exercises	25	38
	Inclusion of high load exercises (> 80% RPM)	5	8
	Correction of thoracic spine posture*	2	3
	Proprioceptive Neuromuscular Facilitation (PNF)*	1	2
Types of exercise equipment	Use of small equipment (e.g. elastic bands)	45	69
	Use of training machines (e.g. pulley, pull-down)	27	42
Setting of exercise treatment	Provision and supervision of supplementary home exercises	42	65
Types of manual techniques	Soft tissue techniques (shoulder or shoulder girdle)	56	86
	Manual mobilisation techniques (shoulder)	51	78
	Manual mobilisation of thoracic spine*	9	14
	Manual mobilisation of ribs*	2	3
	Manual mobilisation of cervical spine*	2	3
Supplementary modalities	Heat or cold applications	14	22
	Therapeutic ultrasound*	1	2

**Recorded in "anything else?" category (physiotherapy report form)*

6.7.4.2 Medical treatment

Table 6.8 presents the summary data of supplementary medical treatment. Of the 28 participants who had additional appointments with Dr Betthäuser during the observation period, i.e. in between the initial presentation and the follow-up, 27 had one appointment, and two had two. The most frequently provided medical treatment was a subacromial corticosteroid injection (25 participants received one injection and three received two); the second most frequent was the application of an elastic tape to the shoulder (11 participants received one application and one received two). One participant was provided with a prescription of oral pain medication. Seven participants received more than one treatment, i.e. both a corticosteroid injection and a tape application.

Table 6.8: Supplementary medical treatment

Medical treatment (n = 65)	**N**	**%**
Additional appointments with Dr Betthäuser	28	43
Any medical treatment	37	57
Subacromial corticosteroid (Triamcinolone) injections	27	42
Elastic tape (shoulder)	12	18
Oral pain medication prescription (Metamizole)	1	2
Sick leave	0	0

6.7.5 Follow-up assessment

The mean (SD) interval between the initial assessment and the completion of the follow-up questionnaires (n = 65) was 97 (17) days, the range 77 to 121 days. 34 participants (52%) completed the follow-up questionnaires at Dr Betthäuser's practice, whereas 31 (48%) completed them at home. The mean (SD) interval between the initial assessment and the follow-up ultrasound scan (n = 52) was 100 (13) days, the range 77-121 days.

6.7.6 Outcomes

6.7.6.1 WORC, GPC and tear progression

The results of the assessment of the primary (WORC_2, WORC_change and $WORC\_change_{ADJ}$) and secondary (GPC, tear progression) study outcomes are presented in *Table 6.9*.

Table 6.9: ***Summary outcome statistics***
Outcomes are listed by type of measurement (continuous/categorical).

Outcome (n)	Measurement	Values		
Continuous outcomes		**x̄**	**SD**	**Range**
Disability (65)	WORC_2 score*	533	427	7-1560
Disability _change (65)	WORC2 – WORC1 score, unadjusted	-363	361	-1248-372
	WORC2 – WORC1 score, RTM-adjusted	-363	341	-1102-387
Categorical outcomes		**N**	**%**	
Perceived Change, GPC (64)	+3	5	8	
	+2	32	50	
	+1	18	28	
	±0	5	8	
	-1	3	5	
	-2	1	2	
	-3	0	0	
Tear progression (PTT FTT) (52)	Yes	2	4	
	No	50	96	

**Includes replaced missing data (see section 6.6.15)*

6.7.6.2 Adverse events

Adverse events were documented by the treating physiotherapists for six participants (9%). All related to temporary exacerbations of the shoulder symptoms, but only one was explicitly linked to the physiotherapy treatment.

6.7.7 Prognostic modelling

6.7.7.1 Ratio of individuals in relation to the number of prognostic factors

As described with the methods, I had planned to include all 10 candidate prognostic factors in the multivariable modelling, and to run one full model with all factors. In fact, I excluded one factor, diabetes, on the grounds of its very low prevalence (n = 4; 6%) in the study population (see *Table 6.5*). In consequence, the analyses were run with a maximum of nine factors. One model (model 6), which was a two-factor model including diabetes and smoking (see *Table 6.1* for the candidate models), was removed from the set of candidate models as, without diabetes, it was no longer a multivariable model. Thus, eight candidate models were analysed. In relation to the full model, the ratio of individuals in relation to the number of prognostic factors was thus approximately 7 (61/9).

6.7.7.2 Analysing the candidate models

All analysed models relate to the 61 complete cases, and their key statistics are presented in *Tables 6.11 (a & b)* to *6.18 (a & b).* The residual plots showed no strong evidence of any violation of assumptions (see *Appendix 6.12*). *Table 6.10* shows how the categorical variables were coded.

Table 6.10: Coding of categorical candidate prognostic factors

Factor	Categories	Code
Sex	Female	1
	Male	0
Physical demands	Yes	1
	No	0
History of shoulder pain	Yes	1
	No	0
Diabetes	Yes	1
	No	0
Smoking	Yes	1
	No	0

Model 1

Table 6.11a: Model 1 summary

Candidate model	N	AIC_C	SEE	R^2_{ADJ}
Age + sex + physical demands + disability ($WORC\_1_{ADJ}$) + pain + history of shoulder pain + symptom duration + smoking + pain catastrophizing (PCS)	9	891	313	0.12

Table 6.11b: Model 1 coefficient statistics

Constant and factors	Unstandardized coefficient B	95% CI	
		Lower limit	Upper limit
Constant	-729	-1315	-144
Age (yr)	5	-3	12
Sex (female or male)	47	-130	224
Physical demands (yes or no)	75	-108	258
Disability ($WORC\_1_{ADJ}$ score)	-0.5	-1	0.03
Pain (mm VAS)	2	-2	6
History of shoulder pain (yes or no)	50	-123	223
Symptom duration (wk)	-0.4	-2	1
Smoking (yes or no)	194	-42	430
Pain catastrophizing (PCS score)	20	9	32

Model 2

Table 6.12a: Model 2 summary

Candidate model	N	AIC_C	SEE	R^2_{ADJ}
Smoking + pain catastrophizing (PCS)	2	880	314	0.11

Table 6.12b: Model 2 coefficient statistics

Constant and factors	Unstandardized coefficient B	95% CI	
		Lower limit	Upper limit
Constant	-599	-766	-433
Smoking (yes, no)	132	-91	355
Pain catastrophizing (PCS score)	14	5	23

Model 3

Table 6.13a: Model 3 summary

Candidate model	N	AIC_C	SEE	R^2_{ADJ}
Age + sex	2	889	336	-0.02

Table 6.13b: Model 3 coefficient statistics

Constant and factors	Unstandardized coefficient B	95% CI	
		Lower limit	Upper limit
Constant	-542	-908	-175
Age (yr)	3	-4	10
Sex (female or male)	37	-143	216

Model 4

Table 6.14a: Model 4 summary

Candidate model	N	AIC_C	SEE	R^2_{ADJ}
Age + sex + physical demands + pain + history of shoulder pain + symptom duration + smoking	7	899	344	-0.06

Table 6.14b: Model 4 coefficient statistics

Constant and factors	Unstandardized coefficient B	95% CI	
		Lower limit	Upper limit
Constant	-727	-1304	-150
Age (yr)	3	-5	12
Sex (female or male)	15	-177	207
Physical demands (yes or no)	40	-160	239
Pain (mm VAS)	2	-1	6
History of shoulder pain (yes or no)	45	-144	234
Symptom duration (wk)	-0.3	-2	2
Smoking (yes or no)	90	-160	341

Model 5

Table 6.15a: Model 5 summary

Candidate model	N	AIC_C	SEE	R^2_{ADJ}
Disability ($WORC\_1_{ADJ}$) + pain catastrophizing (PCS)	2	880	314	0.11

Table 6.15b: Model 5 coefficient statistics

Constant and factors	Unstandardized coefficient B	95% CI	
		Lower limit	Upper limit
Constant	-388	-713	-63
Disability ($WORC\_1_{ADJ}$ score)	-0.3	-1	0.2
Pain catastrophizing (PCS score)	16	6	27

Model 7

(Model 6 was not analysed, see *section 6.7.7.1*)

Table 6.16a: Model 7 summary

Candidate model	N	AIC_C	SEE	R^2_{ADJ}
History of shoulder pain + symptom duration	2	889	336	-0.02

Table 6.16b: Model 7 coefficient statistics

Constant and factors	Unstandardized coefficient B	95% CI	
		Lower limit	Upper limit
Constant	-380	-519	-241
History of shoulder pain (yes or no)	61	-115	236
Symptom duration (wk)	-1	-3	1

Model 8

Table 6.17a: Model 8 summary

Candidate model	N	AIC_C	SEE	R^2_{ADJ}
Pain + history of shoulder pain + symptom duration	3	889	335	-0.01

Table 6.17b: Model 8 coefficient statistics

Constant and factors	Unstandardized coefficient B	95% CI	
		Lower limit	Upper limit
Constant	-513	-776	-249
Pain (mm VAS)	2	-1	5
History of shoulder pain (yes or no)	60	-115	235
Symptom duration (wk)	-1	-2	1

Model 9

Table 6.18a: Model 9 summary

Candidate model	N	AIC_C	SEE	R^2_{ADJ}
Pain + pain catastrophizing (PCS)	2	882	318	0.09

Table 6.18b: Model 9 coefficient statistics

Constant and factors	Unstandardized coefficient B	95% CI	
		Lower limit	Upper limit
Constant	-596	-822	-371
Pain (mm VAS)	1	-3	4
Pain catastrophizing (PCS score)	12	2	22

6.7.7.3 Model comparison

Table 6.19 summarises the key statistics by which the analysed models were compared and ranked (see *section 6.6.17.5*). The models are ordered by size of their AIC_C values (from the lowest to the highest). Where AIC_C values are the same, they are further ordered by magnitude of the SEE. An ΔAIC_C of 0 denotes the model(s) with the lowest AIC_C value representing the best model(s) among all candidate models. The ΔAIC_C values of the other models thus represent their distance to the best model(s).

Table 6.19: Model comparison based on AIC_C differences

Model	N factors	AIC_C	ΔAIC_C	SEE
2	2	880	0	314
5	2	880	0	314
9	2	882	2	318
8	3	889	9	335
3	2	889	9	336
7	2	889	9	336
1	9	891	11	313
4	7	899	19	344

As can be seen in *Table 6.19*, models 2 and 5 had the smallest AIC_C value (AIC_{MIN}) and thus represented the best models. These models were consequently taken as the reference models against which the other models were ranked. The AIC_C of model 9 came third. Its ΔAIC_C was within the range of plausible alternatives to the best models (see *section 6.6.17.5*). All other models had ΔAIC_C values outside this range, thus lacking support. The SEE of model 1 was almost identical to that of the two best models, whereas the other models had larger SEEs. The full model had the highest R^2_{ADJ} (see *Table 6.11a*). The models could at the maximum explain approximately 12% of the variation in the WORC_change.

6.7.7.4 Supplementary analysis: PCS

The aim of the multivariable analysis of the candidate models was to identify at least one promising *multivariable* model. I decided to supplement the multivariable analysis by a univariable linear regression analysis of pain catastrophizing to explore the contribution of this factor alone in predicting the WORC_change[11]. The reason was that the closer inspection of the candidate prognostic factors included in the three models with the lowest AIC_C and ΔAIC_C values (models 2, 5, 9) revealed a noticeable commonality between these models, namely pain catastrophizing (PCS). The summary and coefficient statistics of this univariable model are shown in *Table 6.20 (a & b).* The residual plot can be viewed in *Appendix 6.12*. The residuals were well-behaved.

The analysis was then re-run with the PCS score expressed as 2SD (PCS), to estimate the effect of the difference between a typically high (top third) and typically low (lowest third) PCS value on the $WORC\_change_{ADJ}$. The resulting coefficient statistics are included in *Table 6.20 (a & b).*

[11] The univariable analysis of pain catastrophizing reflects a deviation from protocol; see *Appendix* 6.4.

□

PCS model

Table 6.20a: PCS model summary

Model	N	SEE	R^2_{ADJ}
Pain catastrophizing (PCS)	1	315	0.11

Table 6.20b: PCS and PCS_2SD coefficient statistics

Constant and factors	Unstandardized coefficient B	95% CI	
		Lower limit	Upper limit
Constant	-561	-715	-407
PCS score	13	4	22
PCS_2SD score	232	69	395

6.7.8 Model validation and further analyses

The performance and precision of the analysed candidate models did not justify internal validation. Regarding precision, the models' SEE of between 313 and 344 was higher rather than substantially lower than the MID of 300 derived from the sample data (see *Chapter 7*). The sample size, which was towards the lower end of the target range, precluded useful subgroup analysis.

6.8 DISCUSSION

6.8.1 SUMMARY OF THE STUDY

The prognostic study was designed to develop a prognostic model for the outcome of a phase of conservative treatment with physiotherapy in adults with painful PTTs. Seventy patients were enrolled in a prospective single-group cohort study, which was conducted within a secondary care setting in Hamburg, Germany, between December 2012 and January 2015. Sixty-five participants completed the study. All participants were recruited from an orthopaedic specialist practice. The startpoint was the diagnosis of a painful PTT. All participants underwent conservative treatment including physiotherapy, with or without supplementary medical treatment, over a period of

approximately three months. Ten candidate prognostic factors were assessed at the baseline assessment. The primary outcome measure was the WORC, i.e. the WORC_change from baseline to the follow-up after three to four months, the study's endpoint. Secondary outcome measures were ratings of the perceived change of the shoulder problem on a GPC scale, the incidence of tear progression (PTT to FTT) and the occurrence of physiotherapy-related adverse events. The data of 61 participants were analysed by a multivariable modelling analysis based on the AIC_C approach. Eight predefined candidate models were analysed and compared. In a post-hoc exploratory analysis, one additional univariable model, which included only pain catastrophizing, was explored.

6.8.2 STRUCTURE OF DISCUSSION

The present discussion focuses on the prognostic modelling. (WORC_change, GPC, tear progression and adverse events are discussed in *Chapter 7*). The discussion considers both strengths and limitations of the study. *Textbox 6.2* shows the general structure of this section.

Textbox 6.2: Structure of discussion

Section numbers are provided in brackets.

- Summary of main findings (6.8.3)
- PROBAST self-assessment (risk of bias, applicability and usability) (6.8.4)
- Risk of bias (6.8.5)
  - Participant selection (6.8.5.1)
  - Predictors (i.e. prognostic factors) (6.8.5.2)
  - Outcome (6.8.5.3)
  - Treatment (6.8.5.4)
  - Sample size and participant flow (6.8.5.5)
  - Analysis (6.8.5.6)
- Applicability (6.8.6)
  - Participant selection (6.8.6.1)
  - Predictors (i.e. prognostic factors) (6.8.6.2)
  - Outcome (6.8.6.3)
  - Treatment (6.8.6.4)
- Usability (6.8.7)
- Discussing the findings (6.8.8)
  - Model performance (6.8.8.1)
  - Precision of predictions (6.8.8.2)
  - Prognostic factors (6.8.8.3)
- Comparison with other prognostic model studies (6.8.9)
- Conclusions (6.8.10)
  - Implications for clinical practice (6.8.10.1)
  - Implications for research (6.8.10.2)

6.8.3 SUMMARY OF MAIN FINDINGS

As determined through the AIC_C approach, two out of the eight analysed candidate models, pain catastrophizing and smoking, and pain catastrophizing and disability, were selected as the best models, to which the other models were compared. Only one model, pain catastrophizing and pain, was considered as having substantial support for approximating the best models, i.e. for constituting a plausible alternative to them. The full model had the largest adjusted R^2, which was 0.12. The SEE of the nine candidate models ranged between 313 and 344. A noticeable commonality of the two best models and the model with substantial support for approximating them was that all three included pain catastrophizing. To explore the common factor further, I conducted a complementary univariable regression analysis. The R^2_{ADJ} of this model was 0.11, the SEE 315.

6.8.4 PROBAST ASSESSMENT (RISK OF BIAS, APPLICABILITY AND USABILITY)

I assessed my study with the most recent PROBAST version (20/07/2015, *Appendix 6.13*) that was available at the time of writing of this chapter (January 2016; personal communication with Dr Wolff, 07/01/2016)[12]. Differences between this version and the one I used in the prognostic systematic review (*Chapter 3*) were that two questions (relating to domains 2 and 3) had been removed, and that the wording and layout of the tool had been refined. The coding manual which I had developed for use in the prognostic systematic review (see *Appendix 3.8*) was still applicable. At the time of writing of this chapter there was still no official PROBAST guidance document available (personal communication with Robert Wolff, 11 January 2016).

The PROBAST assessment was based on the full (nine-factor) model. The assessment was independently performed by me and a second rater (Dr Hanchard), and then discussed and agreed. *Table 6.21* shows the summary results of the assessment. A more detailed table, which includes the ratings for all signalling questions, can be viewed in *Appendix 6.14*.

Table 6.21: Summary of PROBAST assessment

	Risk of Bias					Applicability concerns			Overall Judgements		
Domain/ Study	1. Participant Selection	2. Predictors	3. Outcome	4. Sample Size & Flow	5. Analysis	1. Participant Selection	2. Predictors	3. Outcome	Risk of Bias	Applicability	Usability
Braun 2015	+	+	+	+	-	+	+	+	-	+	-

\+ = *low risk/concerns;* - = *high risk/concerns*

[12] The inclusion of this PROBAST version in the thesis appendices was approved by Dr Wolff (personal communication, 17/12/2015).

6.8.5 RISK OF BIAS

As shown in *Table 6.21*, I rated the risk of bias as low for four domains (1, 2, 3 and 4) and as high for one (domain 5). I judged my study overall as at high risk of bias, which was primarily due to issues related to the analysis (domain 5). Most of the signalling questions relating to risk of bias could be answered positively (with “yes”) (see *Appendix 6.14*), which strengthens the internal validity of the study. The design of the study was informed by the most current available methodological guidance on prognosis research as available at the time of planning (see *Chapter 2 section 2.11.1*). This helped me to avoid various potential sources of bias which have also been addressed in the prognostic systematic review (*Chapter 3*).

6.8.5.1 Participant selection

Patients were consecutively enrolled in a prospective cohort study, as recommended for studies developing prognostic models (Steyerberg et al. 2013). Due attention was given to the appropriate inclusion and exclusion of participants. The assessment of patients’ eligibility was exclusively done by Dr Betthäuser, which ensured consistency of the assessments and diagnostic criteria. Patients were excluded in case of any uncertainty about their eligibility. Also, any changes of symptoms or diagnosis that occurred during the observation period were monitored and discussed between Dr Betthäuser and me to ensure that eligibility was maintained throughout the observation period. This minimised the risk of bias by inappropriate inclusions.

The study was designed to enrol participants at a similar state of health. Key aspects, which were the same for all participants, were the presence of shoulder pain and the diagnosis of a PTT at enrolment. It was impossible to determine the precise point in time of the development of the PTT. Similarity of some further aspects related to the baseline state of health, i.e. in particular symptom duration and history of shoulder pain, could not be guaranteed. This similarly applied to baseline measures of pain and disability. Any restrictions of these aspects would have threatened recruitment. Differences, though, were accounted for by candidate prognostic factors. Thus, while similarity of the health status at baseline was as far as possible ensured and partly accounted for by prognostic factors, the study cohort could not be a true inception cohort (Altman 2009).

6.8.5.2 Predictors (i.e. prognostic factors)

I judged the risk of bias associated with the prognostic factors as low. All were carefully defined and assessed in a similar way for all participants. The prospective design of the study guaranteed that the assessment of the factors was blinded to the outcome assessment. All factors are assessable at the intended time of use of the models, i.e. baseline assessment. PROBAST further asks reviewers to rate whether “all relevant predictors are analysed” (signalling question 4 of domain 2). On the grounds of the insufficient scientific knowledge on relevant prognostic factors, I had coded this question as “unanswerable” in the PROBAST coding manual. It was thus not considered for the judgements.

6.8.5.3 Outcome

As discussed in the report on the prognostic systematic review, risk of bias may be introduced through mathematical coupling when one or more prognostic factors are either partly or wholly included in the outcome definition (see *Chapter 3 section 3.5.2*). This problem is relevant to my study, too, because I used the WORC both as outcome measure and as a prognostic factor measure. In my study, though, I accounted for this problem by adjusting the WORC_change for RTM. This adjustment minimised the risk of bias related to mathematical coupling.

All prognostic factors and the outcome were patient-reported. Thus, blinding of the outcome assessment to the predictor information could not be guaranteed. The substantial period of time, though, by which the baseline (prognostic factor) and follow-up (outcome) assessments were separated, would tend to decrease the likelihood that participants remembered the prognostic factor information. Furthermore, as stated with the methods, patients were unaware which of the many parameters assessed at baseline were candidate prognostic factors. Thus, I assumed practical blinding. Empirical evidence on the effects of knowledge of prognostic information on the outcome assessment is yet lacking.

6.8.5.4 Treatment

The treatment of all study participants included physiotherapy, but could further include adjunctive medical treatment such as oral pain medication or local steroid injections. Approximately 40% of participants received a corticosteroid injection, in addition to their physiotherapy treatment. The decision to offer the injection may be associated with (some of) the prognostic factors (e.g. pain intensity, pain duration, level of

catastrophizing), which increases the risk of treatment bias (often referred to as confounding by indication). If the injections were effective, this would have influenced participants' outcome, thereby potentially influencing (bias towards the 0) the association between prognostic factors and outcome, and potentially reducing the predictive performance of the model.

6.8.5.5 Sample size and participant flow

The lack of agreement as to what constitutes an appropriate number of outcome events or individuals in relation to the number of studied factors, and the related difficulty of determining the appropriate sample size for prognostic model studies, have been addressed in the methods part (*section 6.6.14*). The ratio of individuals in relation to the number of prognostic factors for the full model in my study was approximately 7. By this, it was within the range of 5-10 individuals per prognostic factor that I had specified as the minimum number for my study and also for the PROBAST assessments in the prognostic systematic review (*Chapter 3*). Still, the small sample size of the study implies that the representativeness of the sample cannot be assured. The study should consequently be viewed as an exploratory investigation.

The number of losses to follow-up was small (five participants, 7%). Of the 65 participants who completed the study, four (6%) were excluded from the multivariable analysis because of missing prognostic factor data. As explained with the results (*section 6.7.2.3*), I considered the loss of precision of the models by excluding these four cases as minimal.

6.8.5.6 Analysis

I consider my approach to the definition, selection and analysis of the candidate prognostic factors a key strength of my study. This includes the avoidance of any categorisation of continuous factors and of the selection of factors for inclusion in the multivariable analysis based on univariable analysis. These aspects have been discussed in the prognostic systematic review (*Chapter 3*). By using the AIC_C approach, I further avoided any automated selection of prognostic factors within the multivariable modelling analysis[13]. As stated in the introduction to the AIC approach (*section 6.6.17.5*), analysis techniques which rely entirely on statistical criteria, with arbitrary cut-offs for "significance", such as forward, backward and stepwise regression,

[13] The method of selecting prognostic factors within the multivariable analysis is as yet not addressed in PROBAST.

have been linked to biased predictions. In particular, stepwise regression has been claimed to be associated with various problems, including biased high (too large) R^2 values, biased high regression coefficients, biased low (too small) standard errors of regression coefficients estimates, and falsely (too) narrow CIs for the effects and predicted values (Flom & Cassell 2007, Harrell 2001 pp. 56–7).

Information-theoretic approaches such as the AIC approach avoid these problems, thus offering a more rational approach. The advantages of the AIC approach in relation to the definition of candidate models are addressed in the discussion of the applicability. The AIC approach can be conducted using conventional statistical software packages such as SPSS. It is striking that automated factor selection methods are, despite the well documented problems associated with their use, still very popular and widely used in multivariable prognostic modelling studies (Sainani 2013, Steyerberg et al. 2001). Information-theoretic approaches are strikingly less evident in the literature: indeed, I am unaware of any prognostic study in the field of musculoskeletal shoulder complaints in which the AIC approach has been used. In this regard, my use of the approach constitutes an original contribution to knowledge.

The high risk of bias judgement for the PROBAST domain 5 (analysis), which ultimately resulted in the overall high risk of bias judgement, was due to the decision not to further evaluate any of the models in my study because of their low performance and precision. I did not account for model overfitting, nor did I evaluate the models' internal validity. Thus, the validity and reliability of the models is unclear, which puts them at risk of bias.

6.8.6 APPLICABILITY

As shown in *Table 6.21*, I rated concerns about applicability as "low" for all three domains, which led to an overall judgement of "low concerns".

6.8.6.1 Participant selection

The population of interest for this study were adults with painful PTTs. As stated in the background chapter (*Chapter 2 section 2.4*), while the term "painful PTT" is used throughout the thesis for reasons of simplification, the population was operationalised as adults with shoulder pain in the presence of PTTs to acknowledge that the presence of shoulder pain cannot unambiguously be linked to the presence of a PTT. The eligibility criteria of the study were designed to account for this problem, as far as possible, by excluding various other potential sources of shoulder pain. Clearly,

though, this approach limits the applicability of the findings to a highly specific population, which may be the key reason for the significantly smaller number of eligible patients than we had anticipated.

All PTTs were diagnosed by US. US is more specific (94%) than sensitive (68%) to PTTs (Roy et al. 2015). Consequently, while some PTTs might have been missed (false negatives), those identified were almost certainly true positives, so that the study population would have been homogeneous in this respect. As stated in *Chapter 2* (*section 2.2*), the supraspinatus tendon is reportedly the most frequently affected rotator cuff tendon and also usually the first to tear. This is confirmed by my study, as 97% of the participants had an isolated supraspinatus PTT.

An advantage of using US over other imaging modalities such as MRI or MRA is that it allows for a quick and inexpensive assessment and that it is a commonly applied, widely available procedure (*see Chapter 2 section 2.9.3*). A limiting factor related to the applicability of US for diagnosing rotator cuff tears is that it requires sufficient training and experience by the assessor (Lenza et al. 2013). All US assessments within my study were conducted by a certified instructor in ultrasonographic shoulder diagnosis, which strengthens the validity of the diagnoses, but which also means that it is primarily applicable to assessors with similar skills and experience. This also applies to the technical standard of the US equipment.

6.8.6.2 Predictors (i.e. prognostic factors)

Applicability was enhanced by the careful consideration of the candidate prognostic factors and their measurement (see *Chapter 5*). Their wide applicability in most clinical settings and by any health professional was ensured by the fact that all factors are easy to assess, as they do not require any special equipment or skills, and as both PROMs (WORC and PCS) are readily available in various languages. Most of the factors can be assessed in very little time, the only caveat being that the completion (and analysis) of the WORC requires comparably more time than other available shoulder PROMs. This disadvantage, though, appears to be outweighed by its specific tailoring to patients with rotator cuff disorders, its comprehensiveness and the positive ratings of its psychometric properties, as shown in *Chapter 5* (*section 5.5.2.1*).

Using the AIC_C approach enhanced applicability through the required definition of candidate prognostic models. Rather than letting the composition of the model(s) be defined by the application of statistical rules, it was me who defined the candidate models, which allowed for consideration of clinical aspects such as the distinction

between modifiable and non-modifiable factors or the effort required to assess factors (e.g. with or without questionnaires; see *Table 6.1*). Clearly, the nine models I defined represent a selection of possible combinations of factors, and this selection was not mean to be exhaustive. Defining candidate prognostic models by use of theoretical and clinical reasoning avoids a key criticism of automated techniques such as stepwise regression, which is that "it [stepwise regression] allows us not to think about the problem." (Harrell 2001 p. 57). Miles and Shevlin (Miles & Shevlin 2001) have proposed the following analogy to explain the problem of letting the computer decide about the selection of factors within multivariable modelling analysis: "If stepwise regression were used to pack your suitcase, it would select the item of clothing that seemed to be the best – a pair of trousers, for example. Then it would examine which items of clothing fitted, *based on what clothes were already packed.* Underwear does not fit well when trousers are in first, so stepwise regression would reject underwear, as it does not fit the model." (p. 38).

6.8.6.3 Outcome

The applicability of the WORC has been addressed with the prognostic factors, i.e. in the precedent section.

6.8.6.4 Treatment

The approach to the treatment was pragmatic. The physiotherapy treatment followed a broad evidence-based protocol. No restrictions were made on the amount or dose of the physiotherapy. This approach complied both with evidence-based principles and the ethos of an observational study and makes it broadly applicable to conservative treatment including exercise-based physiotherapy of varying content and dose. Similarly, no restriction was applied to the supplementary medical treatment, as supplied by Dr Betthäuser. Clearly, the broad approach may partly explain the observed large variability of the outcome. Although I initially considered undertaking subgroup analyses to explore the potential effects of treatment-related factors (e.g. the amount of physiotherapy or medical treatment), these would have been too small to be informative. Arguably, standardising exercises and manual therapy techniques and defining the dosage of treatment narrower than I did would have helped to ensure uniformity, thus enhancing comparability. However, besides the lack of an evidence-based justification for any such approach, this was not a primary concern of my observational study. Rather, my aim was to achieve a wide applicability across clinical

settings by following standard practice informed by the state of evidence on the interventions of interest, under consideration of the German healthcare regulations.

The findings from my intervention systematic review have been confirmed by more recently published systematic reviews of manual therapy and exercise interventions for impingement related shoulder complaints (Abdulla et al. 2015, Dervey et al. 2014, Desjardins-Charbonneau et al. 2015, Littlewood et al. 2015, Ortega-Castillo & Medina-Porqueres 2016, Wang et al. 2014). Littlewood et al. (2015) reviewed prescription parameters related to therapeutic exercises for patients with rotator cuff tendinopathy. Based on 14 studies, they found some evidence in support of the inclusion of resistance exercises, a higher dose (numbers of sets and repetitions) and a duration of at least three months, but confirmed that the specific optimal parameters are unclear.

6.8.7 USABILITY

None of the candidate models in my study are usable (in clinical practice). The primary reason is that the study was a model development study. The provision of a clinically usable model requires several further stages including external validation and evaluation of clinical impact (*see Chapter 2, section 2.11.2*). Thus, developmental studies, in particular those that have not been validated in any way, *per se* do not provide models that are ready for use in clinical practice (Steyerberg et al. 2013).

The overall very low performance of all candidate models in my study further suggests that none of them can be considered to provide helpful guidance to clinicians. In the light of the lack of knowledge and the difficulties in identifying potentially relevant prognostic factors, as evidenced by the report on the prognostic factor selection process (see *Chapter 5*), the maximum achieved performance (R^2_{ADJ} = 0.12) may be viewed as better than nothing, and a starting point, but it still means that the models are not useful to predict the outcome of a course of conservative treatment with physiotherapy in patients with painful PTTs.

6.8.8 DISCUSSING THE FINDINGS

6.8.8.1 Model performance

A salient finding of the study is that it could not provide a prognostic model with a satisfactory performance. Based on the adjusted R^2, the full model, which had the highest R^2_{ADJ} value, could explain no more than approximately 12% of the variability of

the WORC_change, which means that 88% remained unexplained. This is perhaps not surprising, considering that a vast number of factors could possibly contribute to the prediction of outcomes in individuals, while my investigation was limited to 10 factors. While the 10 factors I investigated performed poorly when combined in various candidate models, they could explain at least some of the variability of the outcome and should thus not be completely discarded based on this study.

The results of my study, though, appear to illustrate a key challenge of developing prognostic models, which is the selection of factors for inclusion in a prognostic model study. The rigorous process by which I selected the ten candidate factors for inclusion in my study (see *Chapter 5*) revealed the dearth of knowledge of relevant factors for predicting outcomes in the field of the conservative treatment of rotator cuff disorders and shoulder pain and the resulting challenge of deciding which factors might potentially be relevant for predicting outcomes in adults with painful PTTs.

As detailed in the chapter on the prognostic factor selection (*Chapter 5*), I included five factors (age, disability, history of shoulder pain, symptom duration, pain) on the grounds that there was reasonably consistent support for their prognostic relevance through clinical evidence and/or expert consensus from either three or more studies or from clinical evidence *and* expert consensus. The results of the multivariable analyses within my study, though, suggest that the prognostic value of these five factors, when analysed in combination and in context with other factors, may be limited. Possible reasons for this finding may include the fact that I did not assess the methodological quality of the included research reports. It has been claimed that many prognostic factor studies are methodologically compromised (Riley et al. 2013). Thus, many studies may be affected by biased outcomes. In this context, the available systematic reviews of prognostic factor studies (Chester et al. 2013, Kuijpers et al. 2004) may have failed to detect relevant sources of risk of bias, as no appropriate instrument for the assessment of risk of bias in prognostic factor studies was available at the time of when they were conducted. Beyond that, I considered any prognostic factor that was considered by the study investigators to have a relevant association with a study outcome, and did not assess the magnitude of the prognostic performance.

6.8.8.2 Precision of predictions

As evidenced by the large SEE and the wide CIs of the regression coefficients, the predictions lack precision, which implies that they are affected by considerable uncertainty and do not provide reliable estimates of population parameters (Hugh 2008). The low precision likely relates to the rather small sample size (also see the

discussion of risk of bias) and variability in the predictors and outcome (Altman & Royston 2000). The low precision adds to the need to interpret the findings with caution and to view this study as exploratory.

6.8.8.3 Prognostic factors

Pain catastrophizing

Pain catastrophizing was the only factor in my study that was part of both the two best models and the third model that was found to constitute a plausible alternative. As the clear intention of my study was to explore and compare prognostic models rather than individual prognostic factors, the complementary univariable analysis and its outcome should be viewed and interpreted very cautiously. It may, though, provide some indication of the relative relevance of pain catastrophizing within the investigated candidate factors.

The potential prognostic relevance of pain catastrophizing is particularly interesting in relation to people with diagnosed PTTs as it seems plausible that patients presenting with shoulder pain who are confronted with a diagnosis of a structural defect such as a PTT may perceive this as threatening. This raises the question of the potential effects of diagnostic imaging on clinical outcomes. A qualitative study (Minns Lowe 2015) which was conducted alongside the United Kingdom Rotator Cuff Surgery (UKUFF) trial tends to support this view as expressed by the following quote from the conference abstract (Results): “Many participants spoke of the importance of diagnosis and impact of visually seeing a rotator cuff tear on scan/screen. Participants considered tears as serious, especially if ‘large’/‘massive’.” However, the observed mean PCS score of 15 in my study indicates that overall, pain catastrophizing was not to a major issue: most (94%) of the participants had a PCS score below the proposed “cut-off score (30) for clinically relevant levels of catastrophizing” (Sullivan 2009 p. 7).

The role of routine diagnostic imaging for assessing musculoskeletal complaints is controversial, though the controversy has mainly centred on (non-specific) low back pain (Flynn 2011, Karel et al. 2015). In this context, the benefits of advanced technological capabilities and the overall increased availability of diagnostic imaging have been challenged by concerns about potential detrimental effects of the routine (or over-) use of diagnostic imaging (in particular MRI), including e.g. misguided clinical decisions based on false-negative diagnoses, increased rates of surgery or stimulation of fear-avoidance and catastrophizing behaviours (Elliott 2011, Flynn 2011, Karel et al.

2015). However, a recent systematic review on the effects of diagnostic imaging in patients with musculoskeletal disorders (Karel et al. 2015) found no relevant studies of shoulder pain populations. The benefits and harms of routine diagnostic imaging in this area of musculoskeletal complaints therefore remain unknown.

The role of pain catastrophizing (same as of psychological or psychosocial factors in general) in predicting outcomes in people with painful rotator cuff disorders is yet insufficiently researched. There is limited evidence from individual heterogeneous studies to indicate that pain catastrophizing may be a relevant factor for predicting outcomes (i.e. symptom intensity, disability, persistence of complaints and sick leave) in people with arm, neck and shoulder complaints (Karels et al. 2007, Larsman et al. 2009) or nonspecific shoulder pain (Kuijpers et al. 2006). In the recent systematic review of primary prognostic factor studies for predicting the outcome of physiotherapy in people with musculoskeletal shoulder pain (including rotator-cuff related disorders) by Chester et al. (2013), though, neither pain catastrophizing nor any other psychosocial factors were identified as showing a consistent association with outcomes in at least two studies that met a predefined set of quality criteria. Limited evidence from individual studies further provides some preliminary indication of the following: a) that pain catastrophizing (among other psychological or psychosocial factors) may have a stronger influence on persistent pain and disability in patients with low back pain compared to patients with shoulder pain (van der Windt et al. 2007); b) that it may be more relevant in chronic compared to acute shoulder pain (Reilingh et al. 2008); and c) that it may be similar in people with non-specific compared to specific diagnoses of complaints of the arm, shoulder and neck (including specific diagnoses such as rotator cuff tendinopathy or tears) (Keijsers et al. 2010).

To my knowledge, the study by Kromer et al. (2014), one of the included studies in my prognostic systematic review (*Chapter 3*), was the first to investigate the role of two psychosocial factors (fear-avoidance beliefs and pain catastrophizing) in a specific population of rotator-cuff related shoulder pain. The authors concluded that neither of these factors was a significant contributor to the prognosis of disability after a three-month period of treatment with physiotherapy. As I judged this study to be affected by a high risk of bias, though, the validity of its findings is to be questioned.

Diabetes

I intended to include all ten candidate prognostic factors in the multivariable analysis, but eventually excluded Diabetes in view of the low number of participants (four) who had Diabetes. The prevalence in the study sample (6%) was below the estimated

overall prevalence of (diagnosed) diabetes of approximately 7% in the German adult population (Diabetes Deutschland 2012). (A more recent estimate is 7-8% (Deutsche Diabetes-Hilfe 2015)). The role of diabetes as a prognostic factor in rotator cuff disorders remains unclear.

6.8.9 COMPARISON WITH OTHER PROGNOSTIC MODEL STUDIES

As evidenced by my prognostic systematic review (*Chapter 3*), this study is the first to investigate prognostic models for predicting the outcome of a period of conservative treatment with physiotherapy in adults with painful PTTs. Thus, no direct comparisons are possible. In the light of the considerable heterogeneity and methodological deficiencies of the available studies exploring prognostic models in adults undergoing conservative treatment with physiotherapy for any type of painful rotator cuff disorders, I considered any further comparison with my study and its findings as uninformative.

6.8.10 CONCLUSIONS

6.8.10.1 Implications for clinical practice

This study is the first to explore prognostic models for the outcome of conservative treatment with physiotherapy in adults with painful PTTs. All analysed candidate models had low performance and precision. The best-performing model could explain only 12% of the variability of the WORC_change. On the grounds of the low performance of the analysed candidate models and the low precision of parameter estimates, none of the models was further evaluated or validated. Given the fact that model development studies *per se* cannot be considered as clinically usable, the study could not provide any model that is ready to be used in clinical practice.

6.8.10.2 Implications for research

Further research is needed to enable provision of a high-performing prognostic model for predicting the outcome of conservative treatment with physiotherapy in adults with painful PTTs. The study was designed in consideration of current methodological guidance for prognosis research and was conducted according to the best possible methodological standards. Furthermore, it is to my knowledge the first prognostic model study in the field of musculoskeletal shoulder complaints in which an information-theoretic analysis approach was applied. The study may thereby provide a

helpful methodological “template” for future studies. It was meticulously reported considering all items required by the TRIPOD statement (Collins et al. 2015, Moons et al. 2015).

The study was limited to the investigation of 10 candidate prognostic factors. Larger studies are needed to enable investigations of larger numbers of factors. This appears particularly relevant given the dearth of knowledge on relevant prognostic factors within the field of shoulder pain and rotator cuff disorders and the implicit challenge of selecting candidate factors for prognostic model studies. Prospective cohort studies are the preferable study design, but prospectively planned analyses of RCT data may also be considered. Collaborative data collection efforts such as multicenter studies, prospective health databases or other data sharing approaches could enhance larger samples. Given the complexity of the design and analysis of prognosis research, the involvement of a statistician with expertise in prognosis research is strongly recommended as it can help to minimize the risk of producing biased predictions.

Further methodological research is needed to determine the most valid and reliable methods for the development of prognostic models. Researchers should carefully design their studies in accordance with current methodological guidance and should keep abreast with the methodological advancements in the field of prognosis research.

6.9 REFERENCES

Abdulla S. Y., Southerst D., Côté P., Shearer H. M., Sutton D., Randhawa K., Varatharajan S., Wong J. J., Yu H., Marchand A.-A., Chrobak K., Woitzik E., Shergill Y., Ferguson B., Stupar M., Nordin M., Jacobs C., Mior S., Carroll L. J., van der Velde G., Taylor-Vaisey A. (2015). Is exercise effective for the management of subacromial impingement syndrome and other soft tissue injuries of the shoulder? A systematic review by the Ontario Protocol for Traffic Injury Management (OPTIMa) Collaboration. *Man Ther*, 20(5), 646-56. doi:10.1016/j.math.2015.03.013.

Altman D. G., Royston P. (2000). What do we mean by validating a prognostic model? *Stat Med*, 19(4), 453–73. doi:10.1002/(SICI)1097-0258(20000229)19:4<453::AID-SIM350>3.0.CO;2-5.

Altman D. G. (2009). Prognostic models: a methodological framework and review of models for breast cancer. *Cancer Invest*, 27(3), 235–43. doi:10.1080/07357900802572110.

Anderson D. A. (2008). *Model based inference in the life sciences. A primer on evidence.* New York: Springer.

Bouwmeester W., Zuithoff N. P. a, Mallett S., Geerlings M. I., Vergouwe Y., Steyerberg E. W., Altman D. G., Moons K. G. M. (2012). Reporting and methods in clinical prediction research: A systematic review. *PLoS Med*, 9(5), 1–12. doi:10.1371/journal.pmed.1001221.

Brozek J. L., Guyatt G. H., Schünemann H. J. (2006). How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health Qual Life Outcomes*, 4, 69. doi:10.1186/1477-7525-4-69.

Bryant L., Shnier R., Bryant C., Murrell G. A. C. (2002). A comparison of clinical estimation, ultrasonography, magnetic resonance imaging, and arthroscopy in determining the size of rotator cuff tears. *J Shoulder Elbow Surg*, 11(3), 219–24. doi:10.1067/mse.2002.121923.

Burnham, K.P.; Anderson D. R. (2002). *Model selection and multimodel inference. A practical information-theoretic approach.* New York: Springer.

Burnham K. P. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociol Methods Res*, 33(2), 261–304. doi:10.1177/0049124104268644.

Burnham K. P., Anderson D. R., Huyvaert K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav Ecol Sociobiol*, 65(1), 23–35. doi:10.1007/s00265-010-1029-6.

Chester R., Shepstone L., Daniell H., Sweeting D., Lewis J., Jerosch-Herold C. (2013). Predicting response to physiotherapy treatment for musculoskeletal shoulder pain: a systematic review. *BMC Musculoskelet Disord*, 14(1), 203. doi:10.1186/1471-2474-14-203.

Collins G. S., Reitsma J. B., Altman D. G., Moons K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*, 350, g7594. doi:10.1136/bmj.g7594.

Dervey E., Marshall S., Rouse S. (2014). Eccentric exercise therapy in the treatment of subacromial impingement syndrome: A critical review. *Int J Ther Rehabil*, 21(7), 338–45.

Desjardins-Charbonneau A., Roy J.-S., Dionne C. E., Frémont P., MacDermid J. C., Desmeules F. (2015). The efficacy of manual therapy for rotator cuff tendinopathy: a systematic review and meta-analysis. *J Orthop Sports Phys Ther*, 45(5), 330–50. doi:10.2519/jospt.2015.5455.

Deutsche Diabetes-Hilfe (2015). *Deutscher Gesundheitsbericht Diabetes 2015*. Mainz. Available at: http://www.diabetesde.org/fileadmin/users/Patientenseite/PDFs_und_TEXTE/Infomaterial/Gesundheitsbericht_2015.pdf [Last accessed 03 June 2016].

Diabetes Deutschland. (2012). *Prävalenz Diabetes (Deutscher Gesundheitssurvey) [Prevalence of diabetes (German health survey].* Available at: http://www.diabetes-deutschland.de/aktuellesituation.html [Last accessed 03 June 2016].

DVSE (2012). Deutsche Vereinigung für Schulter- und Ellbogenchirurgie [German Society for Shoulder and Elbow Surgery] (ed.). Untersuchungstechniken des Schultergelenks. Expertenevaluation auf der Basis einer Literaturanalyse [Assessment techniques of the shoulder joint. Expert evaluation based on an analysis of the literature]. *Obere Extremität*, *7*(Suppl.1), 3–68.

Dworkin R. H., Turk D. C., Farrar J. T., Haythornthwaite J. a, Jensen M. P., Katz N. P., Kerns R. D., Stucki G., Allen R. R., Bellamy N., Carr D. B., Chandler J., Cowan P., Dionne R., Galer B. S., Hertz S., Jadad A. R., Kramer L. D., Manning D. C., Martin S., McCormick C. G., McDermott M. P., McGrath P., Quessy S., Rappaport B. A., Robbins W., Robinson J. P., Rothman M., Royal M. A., Simon L., Stauffer J. W., Stein W., Tollett J., Wernicke J., Witter J. (2005). Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*, 113(1-2), 9–19. doi:10.1016/j.pain.2004.09.012.

Elliott J. M. (2011). The Pearls and Pitfalls of Magnetic Resonance Imaging for the Spine. *J Orthop Sports Phys Ther*, 41(11), 848–60. doi:10.2519/jospt.2011.3636.

Flom, P.L.; Cassell D. L. (2007). Statistics and Data Analysis: Why stepwise and similar selection methods are bad , and what you should use (statistics and data analysis). In *North East SAS Users Group (NESUG) 2007 Annual Conference*. Renaissance Harborplace Hotel, Baltimore, USA: SAS. Available at: http://www.lexjansen.com/pnwsug/2008/DavidCassell-StoppingStepwise.pdf [Last accessed 03 June 2016].

Flynn T. W. (2011). Appropriate Use of Diagnostic Imaging in Low Back Pain - A Reminder That Unnecessary Imaging May Do as Much Harm as Good. *J Orthop Sports Phys Ther*, 41(11), 838–46. doi:10.2519/jospt.2011.3618.

Harrell F. E. (2001). *Regression modeling strategies*. New York: Springer.

□

Health Research Authority (2011). Resources: consent and participant information. Current version available at: http://www.hra.nhs.uk/resources/before-you-apply/consent-and-participation/consent-and-participant-information/ [Last accessed 03 June 2016].

Hedtmann A., Fett H. (1995). Schultersonographie bei Subakromialsyndromen mit Erkrankungen und Verletzungen der Rotatorenmanschette [Sonography of the shoulder in subacromial syndromes with diseases and injuries of the rotator cuff]. *Orthopoäde*, *24*(6), 498–508.

Hedtmann A., Fett A. (2002). Sonographie der Rotatorenmanschette (Ultrasonographic diagnosis of the rotator cuff). *Orthopäde*, *31*, 236–246.

Hopkins W. G. (2002). A new view to statistics: regression to the mean. Available at: http://www.sportsci.org/resource/stats/ [Last accessed 03 June 2016].

Hugh M. L. (2008). Standard error: meaning and interpretation. *Biochem Medica*, 18(1), 7–13.

Iannotti J. P., Ciccone J., Buss D. D., Visotsky J. L., Mascha E., Cotman K., Rawool N. M. (2005). Accuracy of office-based ultrasonography of the shoulder for the diagnosis of rotator cuff tears. *J Bone Joint Surg Am*, 87(6), 1305–11. doi:10.2106/JBJS.D.02100.

Jevsevar D. S., Sanders J., Bozic K. J., Brown G. A. (2015). An Introduction to Clinical Significance in Orthopaedic Outcomes Research. *JBJS Rev*, 3(5), e2. doi:10.2106/JBJS.RVW.N.00064.

Kamper S. J., Maher C. G., Mackay G. (2009). Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther*, 17(3), 163–70.

Kamper S. J., Ostelo R. W. J. G., Knol D. L., Maher C. G., de Vet H. C. W., Hancock M. J. (2010). Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol*, 63(7), 760–6.e1. doi:10.1016/j.jclinepi.2009.09.009.

Karel Y. H. J. M., Verkerk K., Endenburg S., Metselaar S., Verhagen A. P. (2015). Effect of routine diagnostic imaging for patients with musculoskeletal disorders: A meta-analysis. *Eur J Intern Med*, 26(8), 585–95. doi:10.1016/j.ejim.2015.06.018.

Karels C. H., Bierma-Zeinstra S. M. A., Burdorf A., Verhagen A. P., Nauta A. P., Koes B. W. (2007). Social and psychological factors influenced the course of arm, neck and shoulder complaints. *J Clin Epidemiol*, 60(8), 839–48. doi:10.1016/j.jclinepi.2006.11.012.

Keijsers E., Feleus A., Miedema H. S., Koes B. W., Bierma-Zeinstra S. M. A. (2010). Psychosocial factors predicted nonrecovery in both specific and nonspecific diagnoses at arm, neck, and shoulder. *J Clin Epidemiol*, 63(12), 1370–9. doi:10.1016/j.jclinepi.2010.01.015.

Kim Y. S., Heo N. Y., Kim M. W. (2011). The test-retest reliability of supraspinatus cross-sectional area measurement by sonography. *Ann Rehabil Med*, 35(4), 524–8. doi:10.5535/arm.2011.35.4.524.

□

Konermann W., Gruber G. (eds.) (2007). *Ultraschalldiagnostik der Bewegungsorgane. Kursbuch nach den Richtlinien der DEGUM und der DGOU [Ultrasonographic diagnostics of the movement system. Guide based on DEGUM and DGOU standards].* 2nd edn. Stuttgart: Thieme.

Kromer T. O., Sieben J. M., de Bie R. A., Bastiaenen C. H. G. (2014). Influence of fear-avoidance beliefs on disability in patients with subacromial shoulder pain in primary care: a secondary analysis. *Phys Ther*, 94(12), 1775–84. doi:10.2522/ptj.20130587.

Kuijpers T., van Der Windt D. A. W. M., van Der Heijden G. J. M. G., Bouter L. M. (2004). Systematic review of prognostic cohort studies on shoulder disorders. *Pain*, 109(3), 420-31. doi:10.1016/j.pain.2004.02.017.

Kuijpers T., van der Windt D. a W. M., van der Heijden G. J. M. G., Twisk J. W. R., Vergouwe Y., Bouter L. M. (2006). A prediction rule for shoulder pain related sick leave: a prospective cohort study. *BMC Musculoskelet Disord*, 7, 97. doi:10.1186/1471-2474-7-97.

Larsman P., Sandsjö L., Kadefors R., Voerman G., Vollenbroek-Hutten M., Hermens H. (2009). Prognostic factors for intervention effect on neck/shoulder symptom intensity and disability among female computer workers. *J Occup Rehabil*, 19(3), 300–11. doi:10.1007/s10926-009-9186-z.

Lenza M., Buchbinder R., Takwoingi Y., Johnston R. V, Hanchard N. C., Faloppa F. (2013). Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered. *Cochrane Database Syst Rev*, (9), CD009020. doi:10.1002/14651858.CD009020.pub2.

Linden A. (2013). Assessing regression to the mean effects in health care initiatives. *BMC Med Res Methodol*, 13(1), 119. doi:10.1186/1471-2288-13-119.

Littlewood C., Malliaras P., Chance-Larsen K. (2015). Therapeutic exercise for rotator cuff tendinopathy: a systematic review of contextual factors and prescription parameters. *Int J Rehabil Res, 38*(2), 95–106. doi:10.1097/MRR.0000000000000113.

Miles, Jeremy; Shevlin M. (2001). *Applying regression & correlation*. London: Sage Publications.

Minns Lowe C. (2015). The need to prevent conservative treatments being inappropriately compared in clinical research: lessons from the UKUFF trial (a qualitative study). [Conference abstract]. Available at: http://www.physiotherapyuk.org.uk/presentation/need-prevent-conservative-treatments-being-inappropriately-compared-clinical-research [Last accessed 03 June 2016].

Moons K. G. M., Royston P., Vergouwe Y., Grobbee D. E., Altman D. G. (2009). Prognosis and prognostic research: what, why, and how? *BMJ*, 338, b375. doi:10.1136/bmj.b375.

Moons K. G. M., Altman D. G., Reitsma J. B., Ioannidis J. P. a., Macaskill P., Steyerberg E. W., Vickers A. J., Ransohoff D. F., Collins G. S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med*, 162(1), W1. doi:10.7326/M14-0698.

Ortega-Castillo M., Medina-Porqueres I. (2016). Effectiveness of the eccentric exercise therapy in physically active adults with symptomatic shoulder impingement or lateral epicondylar tendinopathy: A systematic review. *J Sci Med Sport*, 19(6), 438–53. doi:10.1016/j.jsams.2015.06.007.

Peduzzi P., Concato J., Kemper E., Holford T. R., Feinstein A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*, 49(12), 1373–9.

Preston C. C., Colman A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychol (Amst),*104(1), 1–15. doi:10.1016/S0001-6918(99)00050-5.

Reilingh M. L., Kuijpers T., Tanja-Harfterkamp A. M., van der Windt D. A. (2008). Course and prognosis of shoulder symptoms in general practice. *Rheumatology (Oxford)*, 47(5), 724–30. doi:10.1093/rheumatology/ken044.

Revicki D., Hays R. D., Cella D., Sloan J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*, 61(2), 102–9. doi:10.1016/j.jclinepi.2007.03.012.

Riley R. D., Hayden J. A., Steyerberg E. W., Moons K. G. M., Abrams K., Kyzas P. A., Malats N., Briggs A., Schroter S., Altman D. G., Hemingway H., PROGRESS Group. (2013). Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*, 10(2), e1001380. doi:10.1371/journal.pmed.1001380.

Roy J.-S., Braen C., Leblond J., Desmeules F., Dionne C. E., MacDermid J. C., Bureau N. J., Fremont P. (2015). Diagnostic accuracy of ultrasonography, MRI and MR arthrography in the characterisation of rotator cuff disorders: a meta-analysis. *Br J Sports Med*, 1–15. doi:10.1136/bjsports-2014-094148.

Royston P., Moons K. G. M., Altman D. G., Vergouwe Y. (2009). Prognosis and prognostic research: Developing a prognostic model. *BMJ*, 338, b604–b604. doi:10.1136/bmj.b604.

Sainani K. L. (2013). Multivariate Regression: The Pitfalls of Automated Variable Selection. *PM R*, 5(9), 791–4. doi:10.1016/j.pmrj.2013.07.007.

Schomacher J. (2008). Gütekriterien der visuellen Analogskala zur Schmerzbewertung (Quality critaria of the visual analogue scale for pain assessment). *Physioscience*, 4(3), 125–33.

Steyerberg E. W., Eijkemans M. J., Harrell F. E., Habbema J. D. (2001). Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making*, 21, 45–56. doi:10.1177/0272989X0102100106.

Steyerberg E. W., Moons K. G. M., van der Windt D. A., Hayden J. A., Perel P., Schroter S., Riley R. D., Hemingway H., Altman D. G. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med*, 10(2), e1001381. doi:10.1371/journal.pmed.1001381.

Steyerberg E. W., Harrell F. E. (2016). Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*, 69, 245–7. doi:10.1016/j.jclinepi.2015.04.005.

Sullivan M. J. L. (2009). *Pain Catastrophizing Scale. User Manual*. Available at: http://sullivan-painresearch.mcgill.ca/pdf/pcs/PCSManual_English.pdf [Last accessed 03 June 2016].

Teefey S. A., Rubin D. A., Middleton W. D., Hildebolt C. F., Leibold R. A., Yamaguchi K. (2004). Detection and quantification of rotator cuff tears. Comparison of ultrasonographic, magnetic resonance imaging, and arthroscopic findings in seventy-one consecutive cases. *J Bone Joint Surg Am*, 86-A(4), 708–16.

Teesside University (2011). *Guidance notes on the preparation of participant information sheets and consent forms & Examples of participant information sheets, consent and assent forms and cover letters.* Internal Teesside University documents (publicly unavailable).

Turner D., Schünemann H. J., Griffith L. E., Beaton D. E., Griffiths A. M., Critch J. N., Guyatt G. H. (2010). The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol*, 63(1), 28–36. doi:10.1016/j.jclinepi.2009.01.024.

van der Windt D. A. W. M., Kuijpers T., Jellema P., van der Heijden G. J. M. G., Bouter L. M. (2007). Do psychological factors predict outcome in both low-back pain and shoulder pain? *Ann Rheum Dis*, 66(3), 313–9. doi:10.1136/ard.2006.053553.

Vittinghoff E., McCulloch C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol*, 165(6), 710–8. doi:10.1093/aje/kwk052.

Wang T. L., Fu B. M., Ngai G., Yung P. (2014). Effect of isokinetic training on shoulder impingement. *Genet Mol Res*, 13(1), 744–757. doi:10.4238/2014.January.31.1.

Wright A., Hannon J., Hegedus E. J., Kavchak A. E. (2012). Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther*, 20(3), 160–6. doi:10.1179/2042618612Y.0000000001.

Orientation Table Chapter 7

Part	Ch.	Title	Aims
ONE	1	General introduction, aims, content and structure of the thesis	1. To provide a general introduction to the topic 2. To summarise the aims, content and structure of the thesis
	2	Background	To provide the relevant topical and conceptual background to the programme of research
TWO	3	Prognostic models in adults undergoing physiotherapy for rotator cuff disorders - a systematic review	To establish the state of evidence on prognostic models in adults undergoing physiotherapy for painful rotator cuff disorders
	4	Developing and validating the physiotherapy protocol for the prognostic study	1. To establish the state of evidence on the effectiveness of physiotherapy interventions for adults with painful atraumatic PTTs 2. To develop and validate the physiotherapy treatment protocol
	5	Selecting and defining the candidate prognostic factors for the prognostic study	1. To identify and select the candidate factors for the prognostic model study 2. To define the specific measures for the selected factors
	6	Predicting the outcome of physiotherapy in adults with painful partial-thickness rotator cuff tears (PTTs) – a prognostic model study	To develop a prognostic model for the outcome of a period of conservative treatment with physiotherapy in adult patients with painful atraumatic PTTs
	7	**Drawing meaningful magnitude-based inferences from the prognostic study – Minimal Important Difference (MID) and responder analysis**	**1. To establish an estimate of the MID of the Western Ontario Rotator Cuff Index (WORC), the primary outcome of the prognostic model study** **2. To apply the estimated MID to an exploratory responder analysis**
THREE	8	Overall summary and conclusions	1. To summarise the research 2. To provide overall conclusions and consider implications
FOUR		Appendices	Appendices to Chapters 3-7

CHAPTER 7
Drawing meaningful magnitude-based inferences from the prognostic study – Minimal Important Difference (MID) and responder analysis

7.1 INTRODUCTION, AIMS AND OBJECTIVES

A secondary aim of my prognostic study (*Chapter 6*) was to enhance interpretation of WORC_change scores. To do so, I estimated the MID of the WORC based on the prognostic study data. The intention of the MID analysis was to enable *magnitude*-based inferences about the WORC_change. I explicitly intended to avoid any inferences based on statistical hypothesis tests, i.e. on the statistical significance of the outcome, because inferences that are grounded on whether a p value (from a statistical null-hypothesis test) is below or above a specific threshold (typically set at 0.05) are widely held to fall short of meaningful interpretation (Batterham & Hopkins 2006, Field 2013, George & Batterham 2015, Shrier & Batterham 2002). The statistical significance of an outcome reflects "the probability of obtaining any value larger than the observed effect (regardless of sign) if the null hypothesis were true" (Batterham & Hopkins 2006 p. 51). However, p values do not provide information about the direction and magnitude of an outcome (Batterham & Hopkins 2006). Also, as statistical significance is affected by sample size and sampling variability, a statistically significant outcome does not necessarily represent a clinically relevant outcome, and conversely, a statistically non-significant outcome does not necessarily represent a clinically non-significant outcome (Batterham & Hopkins 2006, Field 2013). Thresholds for p values (such as 0.05) are further criticised for their arbitrariness (Batterham & Hopkins 2006, Field 2013). Considering these issues, it has been posited that "an overreliance on p values might … lead to unethical errors of interpretation" (Batterham & Hopkins 2006 p. 51).

PROMs like the WORC allow for the comprehensive assessment of those aspects of health status that are relevant to patients (Michener 2011, St-Pierre et al. 2016). PROM change scores quantitatively estimate the direction and magnitude of the change in health status over time; but they do not contextualise this magnitude relative to patients' perception of important change (Guyatt et al. 2002, Schünemann et al. 2006). The consequence is a crucial missing link in interpretation. For example, WORC total scores can range between 0 and 2100 (see *Chapter 5 section 5.5.2.1*). The

sample mean (SD) WORC_change$_{ADJ}$ score in the prognostic study was -363 (341), and the individual outcome scores ranged between -1102 and +387 (*Chapter 6 Table 6.9*). What do these scores mean? How can they be interpreted meaningfully? It would clearly help both investigators and clinicians to have some guidance about what magnitude of a change translates to a clinically important change, i.e. in particular to a change that is important to patients (Brozek et al. 2006, Jevsevar et al. 2015).

Estimates of the MID of an outcome measure offer a threshold for a clinically important outcome in a particular context (Brozek et al. 2006, Guyatt et al. 2002), thereby facilitating responder analysis, "in which a continuous primary efficacy [outcome] measure is dichotomized into 'responders' and 'non-responders'." (Snapinn & Jiang 2007 Abstract). This approach is somewhat crude, however. I therefore complemented the MID analysis by adopting a more rational magnitude-based approach to inferences about the observed WORC_change scores based on the MID estimate. This entailed applying the MID estimate to an exploratory responder analysis of the study participants' individual WORC_change$_{ADJ}$ scores using the method proposed by Hopkins, Batterham et al. (Batterham & Hopkins 2006, Hopkins 2002, 2004, 2007, 2015, Hopkins et al. 2009)[14]. My aim was to illustrate a more sophisticated approach to magni- tude-based inferences that goes beyond the crude dichotomisation of observed outcome values by the MID. To my knowledge, this approach has not yet been applied in the context of shoulder/musculoskeletal outcome MIDs; it thus represents an original contribution to knowledge.

In this chapter, I separately report the MID analysis and the complementary exploratory responder analysis. Accordingly, the chapter is divided into two sub-chapters (*sections 7.2* and *7.3*), each structured into a background, methods, results and discussion section. The discussion sections consider both strengths and limitations of the respective analysis and end with implications for practice and research. Within the discussion of the responder analysis, the other secondary outcomes of the prognostic study, i.e. tear progression and adverse events, are also addressed.

[14] This analysis reflects a deviation from protocol; see *Appendix 6.4*.

7.2 ESTIMATING THE MID OF THE WORC

7.2.1 SPECIFIC BACKGROUND

In 1989, a group of researchers at McMaster University, Canada (Jaeschke et al. 1989), provided a first description of what became known as the “Minimal Clinically Important Difference” (MCID). The MCID was initially defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management” (Jaeschke et al. 1989 p. 408). In 2005, the same group suggested a revised definition: “the smallest difference in score in the outcome of interest that informed patients or proxies perceive as important, either beneficial or harmful, and which would lead the patient or clinician to consider a change in the management” (Schünemann & Guyatt 2005 p. 594). They removed the *C* from MCID, thus terming it MID in order to emphasise the intended focus on “patients’ experience in their day-to-day lives” rather than on the “clinical area” (Schünemann & Guyatt 2005 p. 594).

The MID is still described variously in the literature, and the label M(C)ID encompasses a variety of constructs (Beaton et al. 2002). There are also numerous alternative or related terms, including “Minimum Important Change” (MIC), “Minimal Clinically Important Difference” (MCID), “Minimal Clinically Important Change” (MCIC), “Minimally Practically Important Difference” and “Minimum Worthwile Reduction” (Ferreira et al. 2012 p. 254, George & Batterham 2015 p. 1). However, a review of the M(C)ID literature (Beaton et al. 2002) notes that, “despite different definitions, the common thread is that it is the lower boundary of change that has been defined, in some way, to be important” (p. 110).

Various methodological approaches to the estimation of the MID have been proposed, and the topic is somewhat contentious as knowledge about the best approach is yet lacking (Beaton et al. 2002, Brozek et al. 2006, Jevsevar et al. 2015, Schünemann et al. 2006, Turner et al. 2010, Wright et al. 2012). However, the methodological approaches are commonly broadly categorised into anchor-based and distribution-based approaches (Brozek et al. 2006, Turner et al. 2010, Wright et al. 2012). Anchor-based approaches are generally recommended as the preferable strategy for the MID estimation, but may be combined with distribution-based methods (Revicki et al. 2008, Turner et al. 2010).

Anchor-based approaches estimate the MID by linking the observed outcome to an external criterion, the anchor (Brozek et al. 2006, Jevsevar et al. 2015, Revicki et al. 2008, Wright et al. 2012). This anchor may be a clinical (e.g. range of motion or strength) or patient-reported (e.g. global rating of change) measure, or a combination of both (Revicki et al. 2008, Turner et al. 2010). Patient-reported anchors are generally considered the preferable choice (Revicki et al. 2008, Schünemann & Guyatt 2005). The most widely used patient-based anchor is a Global Perceived Change (GPC) scale (Brozek et al. 2006, Jevsevar et al. 2015, Revicki et al. 2008, Turner et al. 2010, Wright et al. 2012). As described in *Chapter 6* (*section 6.6.9.2*), GPC scales, which may vary in their design and wording, ask patients to rate their perception of the change in their health status related to a particular context over a particular period of time by choosing from various categories that reflect either no change or different levels (e.g. “little” or “much”) of improvement or deterioration (Kamper et al. 2009). It seems intuitive to derive the MID from the smallest (“minimum”) change, but researchers have also used larger changes (e.g. “moderate”) (Turner et al. 2010 p. 29), and the precise number and labels of related GPC categories varies according to the design of the GPC scale used (Turner et al. 2010).

Distribution-based approaches estimate the MID by relying on the statistical properties of the sample such as effect size statistics, the standardized response mean or the standard error of measurement (SEM). They do not apply an external criterion (Brozek et al. 2006, Jevsevar et al. 2015, Revicki et al. 2008, Turner et al. 2010, Wright et al. 2012).

Prior searches of the literature for published estimates of the MID for the WORC (Braun et al. 2013) yielded two estimates determined in studies on populations of adults with impingement-related shoulder pain and by different methods (see further).

7.2.2 METHODS

7.2.2.1 Overall approach and definition of MID

I took an anchor-based approach to estimating the MID for the WORC using the data from the prognostic study: specifically, the data from the 64 participants for whom both the WORC_change and a GPC rating were available. The seven-point GPC scale, one of the secondary outcomes of the prognostic study (see *Chapter 6 section 6.6.9.2* for details), was used as the anchor.

The MID estimate, which was intended to represent a clinically important positive (beneficial) outcome, was established by logistic regression. The GPC was dichotomised into “improved” or “unimproved”. “Improved” comprised +1, +2 or +3 responses on the GPC scale, and thus included all ratings of at least „slightly improved“. “Unimproved” comprised GPC ratings of 0, -1, -2 or -3, thus including all unchanged or deteriorated ratings. The MID was derived from the „improved“ category using a two stage approach. The first stage involved a logistic regression analysis. In the second stage, the probabilities of being improved were calculated for different thresholds of WORC_change.

7.2.2.2 Logistic regression analysis

I conducted a logistic regression analysis with the dichotomised GPC scale (improved/unimproved) as the dependent variable and the WORC_change as the independent variable. As in the prognostic modelling analysis, this was conducted on the RTM-adjusted WORC_change (WORC_change$_{ADJ}$) scores (see *Chapter 6 section 6.6.17.8* for details). The analysis was further adjusted for sex and age, but there was insufficient power to permit presentation of age- and sex-specific MID estimates.

7.2.2.3 Probabilities of being improved for different thresholds of change_WORC

From the regression analysis, I estimated the MID by calculating predicted probabilities of being improved for four WORC_change$_{ADJ}$ thresholds: -100, -200, -300 and -400. The question posed, in other words, was “What is the probability of being ‘improved’ if WORC_change$_{ADJ}$ is -100, -200, -300 and -400 respectively?”. I chose a probability of 90% (a WORC_change$_{ADJ}$ at which nine patients out of 10 would be responders) as a reasonable and intuitive target. In this manner, the WORC_change$_{ADJ}$ threshold with the probability nearest to 90% was designated to represent the MID. I calculated a 95% CI for the probability to provide a measure of variability of the true probability. The analysis was conducted in STATA (version 13.1).

7.2.3 RESULTS

7.2.3.1 Probabilities of being improved for different thresholds of change_WORC

Table 7.1 displays the probabilities of being improved as derived from the logistic regression and calculated for the thresholds of the WORC_change$_{ADJ}$ total scores of -100, -200, -300 and -400. The threshold of -300 for the WORC_change$_{ADJ}$ precisely matched the pre-specified probability of 90%; it was thus accepted as the MID.

Table 7.1: Probabilites of being improved at different WORC_change thresholds
(n = 64)

Level* of WORC_change$_{ADJ}$	Probability	95% CI	
		Lower limit	Upper limit
-100	0.78	0.65	0.92
-200	0.85	0.74	0.96
-300	0.90	0.81	0.99
-400	0.94	0.86	1.01

**relates to margins in STATA*

7.2.3.2 Responders based on the MID

Applying the MID of -300 to classify the 65 study participants as either responders (those whose improvement in the WORC was greater in magnitude than the MID) or non-responders (those whose improvement in the WORC was smaller in magnitude than the MID) resulted in 39 responders (60% of the participants). This is further addressed in the discussion of the responder analysis, in which the implications of different responder definitions, i.e. of different thresholds for being a responder, are compared.

7.2.4 DISCUSSION

7.2.4.1 MID analysis in context

As stated in the background section, both the definition of the MID and the statistical analysis approaches to its determination vary across the research literature, and there is yet no agreement on the best methods. Individual researchers must choose, but it is important to consider that MID estimates are context-dependent: different methods yield different MID estimates (Beaton et al. 2002, Copay et al. 2007, Wright et al. 2012). Also, MID estimates may vary across populations and interventions (Revicki et al. 2008) and also depend on baseline scores (Beaton et al. 2002). It is crucial for researchers to be transparent and explicit about their methods and their specific study contexts. Even so, no single MID estimate will be valid for all applications in which the outcome measure is used (Beaton et al. 2002, Revicki et al. 2008).

7.2.4.2 MID analysis approach

I chose an anchor-based approach for the determination of the MID of the WORC, because, as stated in the background section, anchor-based approaches are generally recommended as the preferable strategy for MID estimation. Moreover, it seemed most reasonable to use an approach in which the WORC_change values were linked to a patient-reported measure of the change of their shoulder problem. A particular strength of this approach is that the probability of a patient being a responder can be calculated at different threshold levels of WORC_change.

Recent reviews of shoulder PROMs which included available MID estimates of various shoulder PROMs suggest that the provision of measures of variability are yet uncommon in shoulder-related MID studies (Roy & Esculier 2011, St-Pierre et al. 2016, Wright & Baumgarten 2010). The approach described here does present CIs, and, moreover, these are easy to interpret: the probability that a patient with a $WORC\_change_{ADJ}$ of -300 was “improved” was 0.81 to 0.99, or put simply, a $WORC\_change_{ADJ}$ of -300 would signify improvement in between 8 out of 10 people and everyone.

7.2.4.3 MID definition

As stated in the background section, GPC scales are the most commonly used anchor to define the threshold of clinical importance in anchor-based MID analyses. I used a

seven-point GPC scale for my MID estimation. The participants' GPC ratings provided estimates of their impression of the overall change of their shoulder problem over time, which I then correlated with the WORC_change$_{ADJ}$ scores in order to provide an estimate of what patients consider a minimal important difference, termed the MID, and to thereby enhance the interpretation of the WORC.

Although GPC scales are the most common anchors, there are some limitations to their use for this purpose. One is related to the fact that GPC ratings need to be translated into an MID definition. This definition is usually made by the researcher (Ferreira et al. 2012). Consequently, while patients provide estimates of how they perceive the change in their complaint, the MID definition ultimately still reflects the researchers' rather than the patients' views (Ferreira et al. 2012). Indeed, this applied to my own analysis, but by accepting a perceived change of at least slightly improved as the mid definition, I aimed for relative robustness. Consequently, MID estimates derived from anchor-based approaches using GPC scales cannot unconditionally be viewed as representing patients' views of the threshold of clinical importance.

As previously outlined (*Chapter 6*), GPC scales have some further limitations. These include the overall yet insufficient knowledge on their validity and reliability, the possible influence of recall bias on patients' ratings (Kamper et al. 2009, Wright et al. 2012) and the concern that GPC scales may not adequately reflect change over time, i.e. that they primarily reflect patients' current health status rather than change (Kamper et al. 2009 & 2010, Wright et al. 2012).

A caveat related to my MID analysis, which relates to the dichotomisation of the GPC ratings into "improved" and "unimproved" in my study (see also further), was that the two categories were unbalanced as most participants (55/64 or 86%) were in the "improved" category, whereas only 9/64 (14%) were in the "unimproved" category, which decreases power (Frazier et al. 2004).

These various limitations and uncertainties should be borne in mind.

7.2.4.4 Comparison with other available MID estimates for the WORC

At the time when I planned this MID analysis two other MID estimates of the WORC were available (Ekeberg et al. 2010, Kirkley et al. 2003a)[15] (see also Braun et al. 2013). To my knowledge, and as confirmed by two recent systematic reviews on the psychometric properties of shoulder questionnaires including the WORC (Huang et al.

[15] The precise origin of the MID analysis presented by Kirkley et al. (2003a), which is a review of various shoulder questionnaires, is unclear as it is not referenced.

2015, St-Pierre et al. 2016), no MID estimates of the WORC have been published since.

The comparison of the approaches to the MID analysis as reported by Ekeberg et al. (2010), Kirkley et al. (2003a) and my own approach provide an illustrative example of the variability of definitions and approaches. A summary of key characteristics and the results of the two studies is presented in *Table 7.2*, which shows that there were both similarities and differences between these studies and my own. Both Kirkley et al. (2003a) and Ekeberg et al. (2010) provide one final MID point estimate. The MID estimate by Ekeberg et al. (2010), though, represents an overall estimate that was established through use of two different approaches.

Table 7.2: Characteristics and results of other WORC MID studies

Study ID	Key characteristics	Analysis approach	GPC and MID definition	MID
Ekeberg 2010	"Rotator cuff disease" (clinical diagnosis, no mention of inclusion/exclusion of tears); n = 121; treatment: corticosteroid injection (local versus systemic); follow-up: after 3 months	Two approaches 1) Mixed anchor-based/distribution-based approach (two estimates); 2) Mean change score of "improved" (3 and 4 points on "main complaint" scale) patients	Initial rating: 18-point "change in main complaint" scale from -9 = "worst possible" to +9 = "best possible" if changed: patients were asked "if they believed that degree of change experienced after treatment of importance to their shoulder condition" (yes/no)? MID: "yes" rating	275
Kirkley 2003	"Chronic cuff tendinitis without tear" (not further specified), n = 44; treatment: subacromial injection (no further information); follow-up: after 6 weeks	Anchor-based	Initial rating: better/worse/same if changed: five-point change scale from 1 = "very little different" to 5 = "a great deal different"; MID: "minimal different" = rating of 1 or 2	245

Study ID = first author, yr

The MID estimate of (-)300 which I established is close to the other two. This estimates' proximity suggests convergent validity, despite differences in the characteristics and methods between the three studies.

On the basis of the known context-dependence as well as the lack of knowledge about the best methodological approach for the determination of the MID, it seems important not to rely on a single estimate but to consider, if possible, a range of MID estimates from various studies and different approaches. Consequently, the MID estimate which I determined should not be viewed as *the* MID and should always be considered within its context, i.e. my study and this specific approach to the MID analysis.

7.2.4.5 Conclusions

Implications for practice

The MID estimate which I established based on the data of my prognostic study provides the first MID estimate of the WORC in adults with painful PTTs who undergo conservative treatment with physiotherapy. It may thus be used by clinicians to enhance the interpretation of WORC_change scores in patients with characteristics that are comparable to those of my study, and within a similar context. Clinicians should be aware, though, of the limitations of MID estimates and should not view the estimate as *the* definitive MID.

Implications for research

Further investigations of MID estimates of the WORC are needed to provide a range of estimates. Further methodological research is need to determine the best methods of determining the MID. In their review of the M(C)ID, (Wright et al. 2012) recommend: "Cautious application of the MCID score both in the clinical setting and in research is prudent until a consensus can be reached on calculation to address the limitations in methodology, population, and baseline." (p. 165). Efforts should be undertaken to increase homogeneity of terminology.

7.3 RESPONDER ANALYSIS

7.3.1 SPECIFIC BACKGROUND

A responder analysis classifies individuals as either responders or non-responders (to treatment) based on the magnitude of their outcome value in relation to a defined threshold of clinical relevance (Jevsevar et al. 2015, Snapinn & Jiang 2007), thereby enabling an easy interpretable measure of benefit. In their guidance on the use of PROMs in clinical trials, the American Food and Drug Administration (FDA) promotes the use of responder analyses as follows (FDA 2009 pp. 24-5): “Regardless of whether the primary endpoint for the clinical trial is based on individual responses to treatment or the group response, it is usually useful to display individual responses, often using an a priori responder definition (i.e., the individual patient PRO [PROM] score change over a predetermined time period that should be interpreted as a treatment benefit). The responder definition is determined empirically and may vary by target population or other clinical trial design characteristics.”

The method for the complementary responder analysis was based on a body of work by Hopkins, Batterham and others (Batterham & Hopkins 2006, Hopkins 2002, 2004, 2007, 2015, Hopkins et al. 2009), who have developed a sophisticated approach to the magnitude-based interpretation of outcomes. The key principle is that inferences should reflect the uncertainty in the true outcome value (Hopkins 2007). Accordingly, the probability that a true outcome value is within a certain level of magnitude takes into account not only the MID, but also the “typical error” of the outcome measurement (see further). Additionally, while the MID is widely used to assign outcomes to one of *two levels of magnitude*, depending on whether their magnitude is below or above the MID, Batterham and Hopkins (2006) have proposed a *three-level scale of magnitude*: “clinically beneficial”, “clinically trivial” and “clinically harmful”. These levels are defined by the positive and negative MID thresholds, which are by default considered equal with opposite signs (Hopkins 2006). *Figure 7.1* illustrates the levels and thresholds (adapted from Hopkins 2006 p. 2). A clinically beneficial outcome reflects any outcome value with a magnitude above the positive MID threshold (i.e. exceeding it in magnitude); a clinically harmful outcome represents any outcome value with a magnitude below the negative MID threshold; and a clinically trivial outcome reflects a outcome value with a magnitude between the positive and negative MID thresholds.

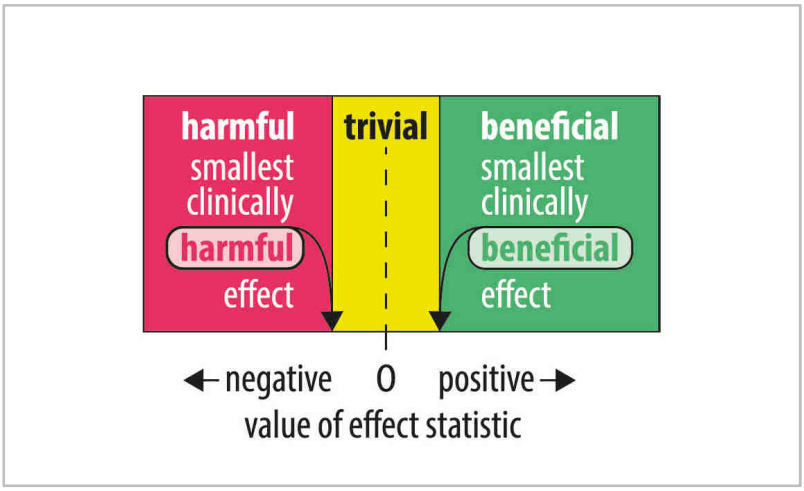


Figure 7.1: Clinically beneficial, trivial and harmful outcomes based on MID thresholds

For individuals' outcome values, the probabilities (from 0.0 to 1.0) can be calculated that the true outcome value is clinically beneficial, trivial or harmful. This is illustrated by *Figure 7.2* (adapted from Hopkins (2006) p. 2). The calculations are based on a t distribution and can be performed in a publicly available, purpose-designed spreadsheet (Hopkins 2007) (see further).

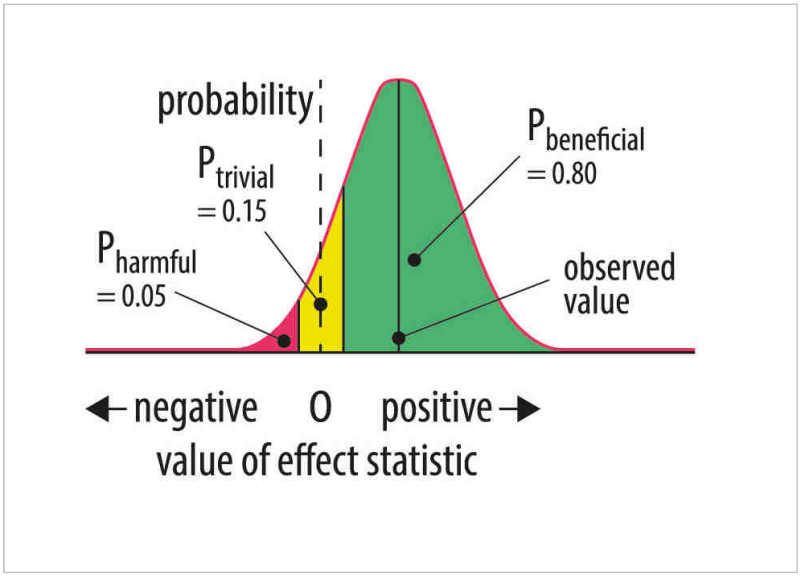


Figure 7.2: Probabilities of a beneficial, trivial or harmful outcome

Hopkins has proposed stratification of the probabilities (or odds) into different levels, each with a qualitative description to enhance meaning (Hopkins 2002). These are shown in *Table 7.3*.

Table 7.3: Levels and qualitative descriptions of probabilities of clinical outcomes

Probability	Odds	Qualitative description
>0.99	>99:1	almost certainly…
0.95-0.99	19:1-99:1	very likely to be…
0.75-0.95	3:1-19:1	likely to be…
0.25-0.75	1:3-3:1	possibly not…
0.05-0.25	1:19-1:3	unlikely to be…
0.01-0.05	1:99-1:19	very unlikely to be…
<0.01	<1:99	not… (almost certainly not…)

As mentioned above, a further feature of the analysis approach is that it considers the typical (or standard) error of the individuals' outcomes. The simple dichotomisation of observed outcome values based on their magnitude in relation to the MID ignores the typical random variation of the measurement, termed the typical error by Hopkins (Hopkins 2004), which expresses the uncertainty in a measurement when it is repeatedly obtained (Hopkins 2004). PROMs, like practically all measurements used in physiotherapy practice and research, have some measurement error, and knowledge of this error helps to distinguish a likely true change from a change that may be primarily due to the error (Revicki et al. 2008). When an individual's outcome value is smaller than the error, it is uncertain whether the change represents a true change or just the error (Beaton 2000). Also, the typical error should ideally be smaller than the MID (Hopkins 2004).

The typical error of a measurement is commonly represented by the SEM, and typically determined in a (test-retest) reliability study, in which the time interval between the repeat measurements must be sufficiently short for the participants not to have changed substantially (Hopkins 2004). In case of studies with longer periods of follow-up, such as my study, Hopkins proposes that the reliability data should reflect the same period of time (Hopkins 2004), meaning that the typical error in such cases represents the typical random variation of the measurement over time (in an untreated group) rather than its test-retest variability. Ideally, the data would be derived from an untreated control group within an RCT. In the absence of such a group within my own study, I could not establish an estimate of the typical error of the WORC from it and thus obtained one from the literature (see further).

The responder analysis presented here should be viewed as illustrative. The aim is to show how the interpretation of outcomes based on the MID can be enhanced by considering the uncertainty in the true values.

7.3.2 METHODS

For the purposes of my illustrative responder analysis, I explored the probabilities of the individuals' true WORC_change$_{ADJ}$ scores being clinically beneficial, i.e. exceeding the positive MID in magnitude, given the typical error of the WORC. The analysis involved two steps: identifying a suitable estimate of the typical error of the WORC and then conducting the responder analysis.

7.3.2.1 Identification of an estimate of the typical error of the WORC

I conducted literature searches in Medline (PubMed), Embase and Cinahl up to week 2, July 2015, for a suitable estimate of the typical error of the WORC. The aim was to identify, if possible, an estimate of the typical error of the WORC from a study with an untreated cohort of patients and comparable characteristics to those of my own study, i.e. in particular with respect to the population (PTTs or impingement-related shoulder pain) and length of follow-up (three to four months). I initially searched for any primary studies of populations of adults with painful rotator cuff disorders in which the WORC was used. Accordingly, I used a pragmatic and broad strategy, in which I combined the terms "Western Ontario Rotator Cuff Index" and "WORC" with terms related to the population of interest (as used in my prognostic systematic review, see *Appendix 3.1 eTable 2*), e.g. "rotator cuff", "shoulder impingement". The searches were supplemented by hand-searches of previously obtained primary studies and systematic reviews in which the WORC was assessed. The findings were then inspected. Relevant studies were required to present repeat measures of the WORC in an untreated cohort of patients. Further, relevant studies had to present either the SEM, the mean WORC_change score with its SD, or the mean WORC pre and post scores with their SDs.

7.3.2.2 Responder analysis

The analysis was based on the WORC_change$_{ADJ}$ values of all study participants (n = 65). The calculations were performed in the purpose-designed spreadsheet by Hopkins (2007) using analysis option 2, "Precision of a change in a measurement based on typical (standard) error of measurement from a reliability study". The spreadsheet is

freely accessible[16]. The calculations require the individuals' baseline and follow-up WORC or WORC_change values, the MID estimate and the following data from a reliability study (or control group): the number of subjects, the number of measurements and the typical error (SEM). Based on these data, the spreadsheet:

- provides individuals' probabilities (also expressed in chances and odds) for a clinically beneficial, trivial and harmful outcome;
- assigns those probabilities to descriptive levels (see *Table 7.3*); and
- provides CIs at different levels of confidence for the individual WORC_change values (see further).

I intended to obtain and present the following data:

- the probabilities (and descriptive levels) of the individuals' true WORC_change$_{ADJ}$ scores being clinically beneficial (i.e. exceeding the MID in magnitude, given the typical error of the WORC);
- the numbers (proportions) of participants at each level of probability related to a beneficial outcome;
- the number (proportion) of participants with a probability of 0.75 of a beneficial outcome, which I used as the ultimate threshold for dichotomising participants to responders and non-responders; and
- the descriptive level of probability of a harmful outcome for all participants whose probability of a beneficial outcome was < 0.10. This was to acknowledge the increasing probability of a harmful outcome with decreasing probability of a beneficial outcome.

In addition to the calculations for the responder analysis, I determined the following measures of variability to describe the uncertainty in the individual and sample true WORC_change$_{ADJ}$ scores (George & Batterham 2015, Hopkins 2015): an SD for the individual responses across the whole sample (SD_{IR}, see further) and CIs for the mean (sample) and individual WORC_change$_{ADJ}$ values.

7.3.2.3 SD for the individual responses across the whole sample

To summarise the individual WORC_change outcomes across the whole sample, and as proposed by Hopkins (2015), I determined an SD (SD_{IR}) by calculating the square root of the difference between the squares of the standard deviations of the

[16] Available at: http://www.sportsci.org/resource/stats/index.html (*Spreadsheets Assessing an individual*) [Last accessed 22 June 2016]

$WORC\_change_{ADJ}$ scores in my study cohort (SD_{STUDY}) and the control group ($SD_{CONTROL}$ from de Witte et al. (2012): $SD_{IR} = \sqrt{(SD_{STUDY}^2 - SD_{CONTROL}^2)}$. This SD can be considered as "the amount by which the net mean effect of the treatment differs typically between individuals" (Hopkins 2015 p. 1444).

7.3.2.4 CIs for the mean (sample) and individual WORC_change values

The 95% CI for the mean (sample) WORC_change was established by a paired-samples t test (between the $WORC1_{ADJ}$ and WORC_2 scores). The 95% CIs for the individuals' WORC_change values were obtained from the spreadsheet (Hopkins 2007).

7.3.3 RESULTS

7.3.3.1 Identification of an estimate of the typical error of the WORC

I identified 18 potentially relevant studies from the searches. A list of these studies, the results of their screening and the references are provided in *Appendix 7.1*. For completeness, the references are also provided in the reference list of this chapter. None of the studies fully complied with the criteria of my own study. In particular, no data was available from an untreated cohort with a comparable length of follow-up. Three studies were unsuitable because the WORC was assessed only once, and ten further studies were unsuitable because there was no untreated cohort. The remaining five studies appeared potentially suitable, the main caveat being that the length of follow-up in these studies, all of which assessed the test-retest reliability of the WORC, was limited to one to two weeks. In three of these studies it was unclear whether or not participants received treatment between the initial assessment and the follow-up. The studies investigated rotator cuff disorders, but none investigated a distinct population of adults with painful PTTs.

I compared my findings to the findings of two recent systematic reviews on the measurement properties of shoulder PROMS for use in populations with rotator cuff disorders (Huang et al. 2015, St-Pierre et al. 2016). Both included comprehensive syntheses of the available psychometric evidence for the WORC. The review by St-Pierre et al. (2016) was the most current and comprehensive one; it included all five reliability studies. No further relevant studies were found. In consequence of these findings, I decided to obtain the most conservative (highest) estimate of the typical error of the WORC from the reliability studies based on the systematic review by

St-Pierre et al. (2016), and to use this as the typical error in my exploratory responder analysis. The caveat was that the data would come from a study that did not fully match the characteristics of my own study and had a shorter follow-up.

The highest available estimate of the typical error of the WORC came from a study by de Witte et al. (2012). In this study, various psychometric properties of the WORC were assessed in 92 patients with rotator cuff disease (35 with rotator cuff tears, 35 with calcific tendinitis and 22 with shoulder impingement), all of whom were considered for participation in one of three independent research projects. The study was conducted in a secondary care setting. The participants' mean age (SD) was 55 (9) years, and 53% were female. Test-retest reliability was established by administering the WORC to the participants twice "within 3 to 14 days" (p. 1616). The questionnaire was first sent to the patients by conventional mail, requesting them to complete it seven to 14 days prior to a scheduled outpatient visit, and was then administered again at the actual visit. It is unclear whether participants received any treatment between the two assessments.

De Witte et al. (2012) provided test-retest reliability data for 83 participants; there was no separate analysis of the three patient subgroups. Test-retest reliability was assessed by calculating the intraclass correlation coefficient (ICC). The SEM was calculated by the following formula: SEM = SD x $\sqrt{}$(1 - ICC) (p. 1613). The resulting SEM for the total WORC was 6.9. As this estimate represented a percentage score of the WORC, I translated it into the absolute score by multiplying it by 21, the resultant score being 144.9. I used this value as the typical error of the WORC in the responder analysis.

7.3.3.2 Responder analysis

Probabilities of being a responder

The result of the assignment of the participants to the different levels of probability based on their individual probabilities are shown in *Table 7.4*. A comprehensive table showing the individual WORC_change$_{ADJ}$ values and probabilities for all 65 participants can be viewed in *Appendix 7.2*. Twenty-five participants (38%) had a probability of ≥ 0.75 of a beneficial outcome and were thus classified as responders (see *section 7.3.2.2* for the responder definition). Twelve participants had a < 0.10 probability of a beneficial outcome. The descriptive level of probability for a harmful outcome was "unlikely" for eight and "possibly" for four (see *Appendix 7.2*).

Table 7.4: Probabilities of a clinically beneficial outcome *(n = 65)*

Probability	Qualitative description	N	%
>0.99	almost certainly	8	12
0.95-0.99	very likely to be	6	9
0.75-0.95	likely to be	11	17
0.25-0.75	possibly	22	34
0.05-0.25	unlikely to be	10	15
0.01-0.05	very unlikely to be	4	6
<0.01	almost certainly not	4	6

SD for individual responses across the sample

The SD_{IR} for the $WORC\_change_{ADJ}$ was 273.

CIs for the mean (sample) and individual WORC_change values

The 95% CI for the sample $WORC\_change_{ADJ}$ ($\bar{x}$ = -363) was -448 to -279. The 95% CIs for the individual $WORC\_change_{ADJ}$ values can be viewed in *Appendix 7.2*.

7.3.4 DISCUSSION

7.3.4.1 Comparison of different responder definitions

As stated in the introduction to this chapter, the aim of conducting the responder analysis was to illustrate an approach to magnitude-based inferences that goes beyond the dichotomisation of outcome values based on the MID. By accounting for the probabilities of individuals' outcomes based on the MID and also the typical error of the outcome measurement, this approach gains in rigor as well as real-world clinical relevance. This may be seen in *Table 7.5*, where it is compared against cruder, but still legitimate, methods to classify responders and non-responders. The availability of a purpose-designed spreadsheet makes the calculations straightforward and should facilitate uptake. Despite this, I am unaware of any published study in the musculoskeletal field in which the approach has been applied. It is important to consider the applied responder definition, in particular when outcomes of responder analyses are compared across studies.

Table 7.5: Summary of different approaches to the interpretation of the WORC_change

Approach (n)	Responder definition	Improved	
		N	%
Any improvement in WORC (65)	Negative WORC_change score (WORC2 < WORC1)	55	85
Improvement according to GPC (64)	GPC rating of +1, +2 or +3	55	86
Improvement exceeding the typical error (65)	WORC_change score exceeding -144.9	49	75
Improvement exceeding the MID (65)	WORC_change score exceeding -300	39	60
Improvement based on probabilities of beneficial outcome (65)	□0.75 (likely, very likely or almost certainly beneficial)	25	38

Considering that the responder analysis relates to the outcome of a single-group cohort study, it was not my intention to make a statement about the effectiveness of the intervention. A strong caveat is that the estimate of the typical error of the WORC was taken from a different study with limited comparability to my own study, including, in particular, a shorter follow-up. The magnitude of the variability of the WORC in an untreated cohort of patients with painful PTTs (matching the criteria of my study) over a period of three to four months is unknown, but is likely that it is larger than the value used.

Furthermore, the disadvantages of dichotomising continuous variables (such as loss of information and power), which have been addressed in the context of the prognostic systematic review and the prognostic study, likewise apply to responder analyses and need to be borne in mind (Jevsevar et al. 2015, Snapinn & Jiang 2007).

7.3.4.2 Balancing benefits against potential downsides or harms

The responder definition which I suggested for this exploratory analysis was based on a □ 0.75 probability of a beneficial outcome. Thus, it did not further consider the probabilities of a harmful outcome. The presented levels of probabilities of harmful outcome for those participants who had a < 0.10 probability of a beneficial outcome provides evidence that the probability of a harmful outcome was very low in the sample. Without doubt, the response to an intervention needs to be viewed in the

context of potential downsides such as the costs or any inconveniences that may be associated with the intervention (Schünemann et al. 2006). As described in *Chapter 4* (*section 4.3.3*), apart from temporary exacerbations of symptoms, no known risks or serious side effects are reported in the literature on the physiotherapy treatment (manual therapy and exercises) of adults with painful rotator cuff disorders. The documentation of adverse events in my study showed that a small number of participants experienced temporary exacerbations of symptoms, but besides, no other adverse events were reported. This indicates that the treatment was overall tolerated well by the participants. Tear progression was another secondary outcome associated with the potential downsides of conservative treatment with physiotherapy. Again, though, the number of participants whose PTT progressed to an FTT over the observation period was very low. Whether the observed progression was related to the treatment or not is unclear.

7.3.4.3 Conclusions

The exploratory responder analysis presented in this part chapter illustrates a sophisticated, robust approach to magnitude-based inferences about outcomes based on the uncertainty in the true outcome value, which enables a realistic interpretation of responders to an intervention. Clinicians and researchers need to be aware that reported rates of responders often fall short of factoring in the typical error of the measurement and of considering the probabilities of being responders, and thus likely overestimate the rates of responders.

7.4 REFERENCES

Baskurt Z., Baskurt F., Gelecek N., Özkan M. H. (2011). The effectiveness of scapular stabilization exercise in the patients with subacromial impingement syndrome. *J Back Musculoskelet Rehabil*, 24(3), 173-9. doi:10.3233/BMR-2011-0291.

Batterham A. M., Hopkins W. G. (2006). Making meaningful inferences about magnitudes. *Int J Sports Physiol Perform*, 1(1), 50-7.

Beaton D. E. (2000). Understanding the relevance of measured change through studies of responsiveness. *Spine (Phila Pa 1976),* 25(24), 3192-9.

Beaton D. E., Boers M., Wells G. A. (2002). Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol*, 14(2), 109-14. doi:10.1097/00002281-200203000-00006.

Bernhardsson S., Klintberg I. H., Wendt G. K. (2011). Evaluation of an exercise concept focusing on eccentric strength training of the rotator cuff for patients with subacromial impingement syndrome. *Clin Rehabil*, 25(1), 69-78. doi:10.1177/0269215510376005.

Braun C., Bularczyk M., Heintsch J., Hanchard N. C. A. (2013). Manual therapy and exercises for shoulder impingement revisited. *Phys Ther Rev*, 18(4), 263-84. doi:10.1179/108331913X13709388114510.

Brozek J. L., Guyatt G. H., Schünemann H. J. (2006). How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health Qual Life Outcomes*, 4, 69. doi:10.1186/1477-7525-4-69.

Copay A. G., Subach B. R., Glassman S. D., Polly D. W., Schuler T. C. (2007). Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J*, 7(5), 541-6. doi:10.1016/j.spinee.2007.01.008.

de Witte P. B., Henseler J. F., Nagels J., Vliet Vlieland T. P. M., Nelissen R. G. H. H. (2012). The Western Ontario rotator cuff index in rotator cuff disease patients: a comprehensive reliability and responsiveness validation study. *Am J Sports Med*, 40(7), 1611-9. doi:10.1177/0363546512446591.

Dilek B., Gulbahar S., Gundogdu M., Ergin B., Manisali M., Ozkan M., Akalin E. (2016). Efficacy of Proprioceptive Exercises in Patients with Subacromial Impingement Syndrome: A Single-Blinded Randomized Controlled Study. *Am J Phys Med Rehabil*, 95(3), 169-82. doi:10.1097/PHM.0000000000000327.

Ekeberg O. M., Bautz-Holter E., Keller A., Tveitå E. K., Juel N. G., Brox J. I. (2010). A questionnaire found disease-specific WORC index is not more responsive than SPADI and OSS in rotator cuff disease. *J Clin Epidemiol*, 63(5), 575-84. doi:10.1016/j.jclinepi.2009.07.012.

Ekeberg O. M., Bautz-Holter E., Tveitå E. K., Keller A., Juel N. G., Brox J. I. (2008). Agreement, reliability and validity in 3 shoulder questionnaires in patients with rotator cuff disease. *BMC Musculoskelet Disord*, 9, 68. doi:10.1186/1471-2474-9-68.

FDA (2009). *Guidance for Industry. Patient-reported outcome measures: Use in medical product development to support labeling claims.* Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf [Last accessed 03 June 2016].

Ferreira M. L., Herbert R. D., Ferreira P. H., Latimer J., Ostelo R. W., Nascimento D. P., Smeets R. J. (2012). A critical review of methods used to determine the smallest worthwhile effect of interventions for low back pain. *J Clin Epidemiol*, 65(3), 253-61. doi:10.1016/j.jclinepi.2011.06.018.

Field A. (2013). *Discovering statistics using SPSS*. 4th edn. London: Sage Publications.

Frazier P. A., Tix A. P., Barron K. E. (2004). Testing Moderator and Mediator Effects in Counseling Psychology Research. *J Couns Psychol*, 51(1), 2004. doi:10.1037/0022-0167.51.1.115.

George K., Batterham A. M. (2015). So what does this all mean? *Phys Ther Sport*, 16(1), 1-2. doi:10.1016/j.ptsp.2014.10.005.

Guyatt G. H., Osoba D., Wu A. W., Wyrwich K. W., Norman G. R. (2002). Methods to explain the clinical significance of health status measures. *Mayo Clin. Proc.*, 77(4), 371-83.

Haik M. N., Alburquerque-Sendín F., Silva C. Z., Siqueira-Junior A. L., Ribeiro I. L., Camargo P. R. (2014). Scapular kinematics pre- and post-thoracic thrust manipulation in individuals with and without shoulder impingement symptoms: a randomized controlled study. *J Orthop Sports Phys Ther*, 44(7), 475-87. doi:10.2519/jospt.2014.4760.

Holtby R., Razmjou H. (2005). Measurement properties of the Western Ontario rotator cuff outcome measure: a preliminary report. *J Shoulder Elbow Surg*, 14(5), 506-10. doi:10.1016/j.jse.2005.02.017.

Hopkins W. G. (2002). Probabilities of Clinical or Practical Significance. *Sportscience*, 6. Available at: http://www.sportsci.org/jour/0201/wghprob.htm [Last accessed 03 June 2016].

Hopkins W. G. (2004). How to Interpret Changes in an Athletic Performance Test. *Sportscience,* 8, 1-7. Available at: http://www.sportsci.org/jour/04/wghtests.pdf [Last accessed 03 June 2016].

Hopkins W. G. (2006). Making Inferences: Clinical vs Statistical Significance [Power point presentation]. Available at: http://www.sportsci.org/2006/Making_inferences.pdf [Last accessed 03 June 2016].

Hopkins W. G. (2007). A Spreadsheet for Deriving a Confidence Interval, Mechanistic Inference and Clinical Inference from a P Value. *Sportscience*, *11*, 16-20. Spreadsheet vailable at: http://www.sportsci.org/resource/stats/index.html (*Spreadsheets Assessing an individual*) [Last accessed 22 June 2016].

Hopkins W. G., Marshall S. W., Batterham A. M., Hanin J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc*, 41(1), 3-13. doi:10.1249/MSS.0b013e31818cb278.

Hopkins W. G. (2015). Individual responses made easy. *J Appl Physiol*, 118(12), 1444-6. doi:10.1152/japplphysiol.00098.2015.

Huang H., Grant J. A., Miller B. S., Mirza F. M., Gagnier J. J. (2015). A Systematic Review of the Psychometric Properties of Patient-Reported Outcome Instruments for Use in Patients With Rotator Cuff Disease. *Am J Sports Med*, 43(10), 2572-82. doi:10.1177/0363546514565096.

Jaeschke R., Singer J., Guyatt G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*, 10(4), 407-15.

Jevsevar D. S., Sanders J., Bozic K. J., Brown G. A. (2015). An Introduction to Clinical Significance in Orthopaedic Outcomes Research. *JBJS Rev*, 3(5), e2. doi:10.2106/JBJS.RVW.N.00064.

Kamper S. J., Maher C. G., Mackay G. (2009). Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther*, 17(3), 163-70.

Kamper S. J., Ostelo R. W. J. G., Knol D. L., Maher C. G., de Vet H. C. W., Hancock M. J. (2010). Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol*, 63(7), 760-6.e1. doi:10.1016/j.jclinepi.2009.09.009.

Kirkley A., Griffin S., Dainty K. (2003a). Scoring systems for the functional assessment of the shoulder. *Arthroscopy*, 19(10), 1109-20. doi:10.1016/j.arthro.2003.10.030.

Kirkley A., Alvarez C., Griffin S. (2003b). The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. *Clin J Sport Med*, 13(2), 84-92.

Kuhn J. E., Dunn W. R., Sanders R., An Q., Baumgarten K. M., Bishop J. Y., Brophy R. H., Carey J. L., Holloway B. G., Jones G. L., Ma C. B., Marx R. G., McCarty E. C., Poddar S. K., Smith M. V, Spencer E. E., Vidal A. F., Wolf B. R., Wright R. W., MOON Shoulder Group. (2013). Effectiveness of physical therapy in treating atraumatic full-thickness rotator cuff tears: a multicenter prospective cohort study. *J Shoulder Elbow Surg*, 22(10), 1371-9. doi:10.1016/j.jse.2013.01.026.

Lopes A. D., Ciconelli R. M., Carrera E. F., Griffin S., Faloppa F., Dos Reis F. B. (2008). Validity and reliability of the Western Ontario Rotator Cuff Index (WORC) for use in Brazil. *Clin J Sport Med*, 18(3), 266-72. doi:10.1097/JSM.0b013e31817282f4.

Martins L. V, Marziale M. H. P. (2012). Assessment of proprioceptive exercises in the treatment of rotator cuff disorders in nursing professionals: a randomized controlled clinical trial. *Rev Bras Fisioter*, 16(6), 502-9.

Michener L. A. (2011). Patient- and clinician-rated outcome measures for clinical decision making in rehabilitation. *J. Sport Rehabil.*, *20*(1), 37-45.

Provencher M. T., Frank R. M., Macian D., Dewing C. B., Ghodadra N. S., Carney J., LeClere L., Solomon D. J. (2012). An analysis of shoulder outcomes scores in 275 consecutive patients: disease-specific correlation across multiple shoulder conditions. *Mil Med*, 177(8), 975-82.

Razmjou H., Bean A., van Osnabrugge V., MacDermid J. C., Holtby R. (2006). Cross-sectional and longitudinal construct validity of two rotator cuff disease-specific outcome measures. *BMC Musculoskelet Disord*, 7, 26. doi:10.1186/1471-2474-7-26.

Razmjou H., Bean A., Macdermid J. C., van Osnabrugge V., Travers N., Holtby R. (2008). Convergent validity of the constant-murley outcome measure in patients with rotator cuff disease. *Physiother Canada*, 60(1), 72-9. doi:10.3138/physio/60/1/72.

Revicki D., Hays R. D., Cella D., Sloan J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*, 61(2), 102-9. doi:10.1016/j.jclinepi.2007.03.012.

Roy J.-S., Esculier J.-F. (2011). Psychometric evidence for clinical outcome measures assessing shoulder disorders. *Phys Ther Rev*, 16(5), 331-46. doi:10.1179/1743288X11Y.0000000043.

Schünemann H. J., Guyatt G. H. (2005). Commentary - goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res*, 40(2), 593-7. doi:10.1111/j.1475-6773.2005.00374.x.

Schünemann H. J., Akl E. A., Guyatt G. H. (2006). Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. *Health Qual Life Outcomes*, 4, 62. doi:10.1186/1477-7525-4-62.

Shrier I., Batterham A. M. (2002). Clinically relevant? *Clin J Sport Med*, 12(6), 328-30.

Snapinn S. M., Jiang Q. (2007). Responder analyses and the assessment of a clinically relevant treatment effect. *Trials*, 8(1), 31. doi:10.1186/1745-6215-8-31.

St-Pierre C., Desmeules F., Dionne C. E., Frémont P., MacDermid J. C., Roy J.-S. (2016). Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. *Disabil Rehabil*, 38(2), 103-22. doi:10.3109/09638288.2015.1027004.

Subasi, V.; Toktas, H.; Demirdal, Ü.S.; Türel, A.; Cakir, T.; Kavuncu V. (2012). Water-based versus land-based exercise program for the management of shoulder impingement syndrome. *Turk J Phys Med Rehab*, 58(2), 79-84.

Turner D., Schünemann H. J., Griffith L. E., Beaton D. E., Griffiths A. M., Critch J. N., Guyatt G. H. (2010). The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol*, 63(1), 28-36. doi:10.1016/j.jclinepi.2009.01.024.

Wessel J., Razmjou H., Mewa Y., Holtby R. (2005). The factor validity of the Western Ontario Rotator Cuff Index. *BMC Musculoskelet Disord*, 6, 22. doi:10.1186/1471-2474-6-22.

Wiertsema S. H., Rietberg M. B., Hekman K. M., Schothorst M., Steultjens M. P., Dekker J. (2013). Reproducibility of the Dutch version of the Western Ontario rotator cuff Index. *J Shoulder Elbow Surg*, 22(2), 165-70. doi:10.1016/j.jse.2012.03.014.

Wright A., Hannon J., Hegedus E. J., Kavchak A. E. (2012). Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther*, 20, 160-6. doi:10.1179/2042618612Y.0000000001.

Wright R. W., Baumgarten K. M. (2010). Shoulder outcomes measures. *J Am Acad Orthop Surg*, 18(7), 436-44.

PART THREE:

OVERALL SUMMARY AND CONCLUSIONS

Orientation Table Chapter 8

Part	Ch.	Title	Aims
ONE	1	General introduction, aims, content and structure of the thesis	1. To provide a general introduction to the topic 2. To summarise the aims, content and structure of the thesis
	2	Background	To provide the relevant topical and conceptual background to the programme of research
TWO	3	Prognostic models in adults undergoing physiotherapy for rotator cuff disorders - a systematic review	To establish the state of evidence on prognostic models in adults undergoing physiotherapy for painful rotator cuff disorders
	4	Developing and validating the physiotherapy protocol for the prognostic study	1. To establish the state of evidence on the effectiveness of physiotherapy interventions for adults with painful atraumatic PTTs 2. To develop and validate the physiotherapy treatment protocol
	5	Selecting and defining the candidate prognostic factors for the prognostic study	1. To identify and select the candidate factors for the prognostic model study 2. To define the specific measures for the selected factors
	6	Predicting the outcome of physiotherapy in adults with painful partial-thickness rotator cuff tears (PTTs) – a prognostic model study	To develop a prognostic model for the outcome of a period of conservative treatment with physiotherapy in adult patients with painful atraumatic PTTs
	7	Drawing meaningful magnitude-based inferences from the prognostic study – Minimal Important Difference (MID) and responder analysis	1. To establish an estimate of the MID of the Western Ontario Rotator Cuff Index (WORC), the primary outcome of the prognostic model study 2. To apply the estimated MID to an exploratory responder analysis
THREE	**8**	**Overall summary and conclusions**	**1. To summarise the research** **2. To provide overall conclusions and consider implications**
FOUR		Appendices	Appendices to Chapters 3-7

CHAPTER 8
Overall summary and conclusions

8.1 OVERALL SUMMARY OF THE RESEARCH

The programme of research presented in this thesis was designed to explore the role of conservative treatment in the management of people with painful PTTs. The primary aim was to develop a prognostic model for the outcome of physiotherapy in adults with painful PTTs. This was addressed by a prospective prognostic model study (*Chapter 6*), the need for which was underpinned by a systematic review of studies exploring prognostic models for predicting outcomes in adults undergoing conservative treatment with physiotherapy for rotator cuff disorders (*Chapter 3*). Five heterogeneous studies, of which none addressed a distinct population of PTTs, were included in this review. All of these studies were judged as at high risk of bias, and none of the analysed models was found to be usable in practice.

The prognostic study was informed by the most current available methodological guidance on prognosis research (see *Chapter 2*), which helped to prevent various potential sources of bias. The physiotherapy treatment protocol for the study was based on two systematic reviews on the effectiveness of manual therapy and exercises for impingement-related shoulder pain (*Chapter 4*). These reviews, which included eight systematic reviews and 15 subsequent RCTs, showed that, while there was evidence that exercises with or without manual therapy are effective in improving clinical outcomes in these patients, it was not possible to conclude on the optimal treatment parameters such as the type and dose of manual therapy or exercises. In view of these findings, the physiotherapy protocol was kept broad. It was piloted and validated by several physiotherapists.

The selection of 10 candidate prognostic factors for inclusion in the study followed a rigorous process by which I aimed to give preference to factors for which there was supporting evidence for their relevance to the study question (*Chapter 5*). Selection involved comprehensive searches of the literature and considered clinical evidence, expert consensus and peer consensus. The searches focussed on painful rotator cuff disorder populations, but evidence from general shoulder pain populations was also considered for systematic reviews and expert consensus studies. Inclusion was limited to factors elicited at the baseline assessment. Selection further considered the relevance of the factors to the study population, the properties of their available

measurements and their applicability/practicability in clinical practice. Twenty-three primary studies, a systematic review of prognostic studies and a study involving a Delphi (expert) consensus process were identified from the searches and informed the selection, providing overall limited guidance. The evidence base for most factors was very weak. The factors were very heterogeneous, as were the approaches used for their measurement. Age, disability and symptom duration were the only factors with reasonably consistent evidence of prognostic value pertaining to clinical outcomes of conservative treatment in patients with rotator cuff disorders. Thereby, the prognostic relevance of most of the candidate factors in my study was unclear. The 10 factors which I investigated were: age, sex, physical demands, disability, pain, history of shoulder pain, symptom duration, diabetes, smoking and pain catastrophizing.

The prognostic model study was conducted within a secondary care setting in Hamburg, Germany. Following initial assessment by an orthopaedic medial specialist, participants were followed over a three-month period of physiotherapy, with or without adjunctive medical treatment. Follow-up was after three to four months. Sixty-five participants completed the study, and the data from 61 participants were included in the multivariable modelling analysis. Eight pre-defined candidate models were analysed by multivariable regression and by applying an information-theoretic approach (AIC_C). The resulting best models were: pain catastrophizing (PCS) and smoking, and disability ($WORC\_1_{ADJ}$) and pain catastrophizing (PCS). One further model was found to represent a plausible alternative: pain and pain catastrophizing (PCS). All models had a poor performance (as shown by their R^2_{ADJ} values) and precision (as shown by their SEE). As pain catastrophizing was part of all best models, it was analysed by a complementary univariable analysis, which showed a similar performance as the multivariable models. Given the low performance and precision of the analysed candidate models, any further analyses or internal validation were unjustified. The relatively small sample size precluded any informative subgroup analysis and implies that the study should be viewed as exploratory.

Complementary to the prognostic study, I aimed to enhance interpretation of the observed WORC_change (i.e. primary outcome) scores by estimating the MID of the WORC. The estimate was determined through an anchor-based approach and using logistic regression. A GPC scale was used as the anchor. The estimated MID was -300 (at the 90% level of probability). A caveat of the analysis was that the two categories (“improved”/”unimproved”) into which I dichotomised the GPC scale were unbalanced as most participants were in the “improved” category.

I applied the MID to a more sophisticated exploratory responder analysis based on the probabilities of the participants' true WORC_change$_{ADJ}$ values being clinically beneficial, i.e. exceeding the (positive) MID in magnitude, given the typical error of the WORC. I determined an estimate of the typical error from the literature. Based on a responder definition of a 0.75 probability of having a beneficial outcome, 38% of the participants were responders. I contrasted this result with several other responder definitions. I further established several measures of variability to describe the uncertainty in the individual and sample true WORC_change$_{ADJ}$ scores. A caveat of the responder analysis was that the typical error of the WORC was obtained from a different study with a shorter follow-up.

I acknowledged the importance of balancing beneficial outcomes against potential downsides or harms by attention to the secondary study outcomes. In terms of benefits, 86% of the participants rated their shoulder problem on the GPC scale as "improved", and 58% rated it as either "much better" or "completely recovered". In terms of harms, the incidence of tear progression (to an FTT) at follow-up was 4%, and physiotherapy-related adverse events, which were reported for 9% of the participants, were limited to temporary exacerbations of the shoulder symptoms. These findings provide some preliminary indication that conservative treatment with physiotherapy may be both beneficial and safe for most adults with painful PTTs, the caveat being that the findings relate to data from an uncontrolled cohort study, which means that they do not reflect the *effectiveness* of the provided treatment.

8.2 OVERALL SUMMARY OF ORIGINAL CONTRIBUTIONS TO KNOWLEDGE

The originality of the contribution to knowledge provided by the programme of research presented in this thesis is evidenced by the following key components, namely:

- the first systematic review on prognostic models for predicting outcomes in adults undergoing conservative treatment with physiotherapy for painful rotator cuff disorders;
- the first prognostic model study exploring prognostic models for the outcome of conservative treatment with physiotherapy in adults with painful PTTs;
- the first analysis of an estimate of the MID of the WORC in a population of adults with painful PTTs; and

- the first application of a responder analysis based on the probabilities of a beneficial outcome beyond the MID given the typical error of the WORC in a population of adults with a painful shoulder disorder.

These key components are complemented by the novel application of various methodological aspects as described in the individual reports.

8.3 OVERALL CONCLUSIONS

Traditionally, shoulder pain has been classified by clinical diagnosis for research (as in the present prognosis study) and clinical decision-making; but there have been criticisms of this model, clearly crystallised by Schellingerhout et al. (2008). They argued that the diagnostic labeling model was failed, largely because of non-uniform diagnostic criteria, and proposed a general “shoulder pain” label, which might eventually be sub-grouped according to prognosis. Superficially, my study, which was grounded in a specific diagnosis, may appear at odds with their proposals; but closer scrutiny reveals broad compatibility and common purpose. First, the diagnosis in my study was not merely a label, but true - due to the high specificity of diagnostic ultrasound for detecting PTTs, this diagnosis could confidently be ascribed to the whole sample. Second, though within a known sub-population rather than a general shoulder pain population, I expressly set out to investigate prognostic models that might ultimately help to inform useful clinical subgrouping.

8.3.1 Implications for practice

As yet, no clinically usable multivariable model of baseline factors is available for predicting the outcome of conservative treatment with physiotherapy in adults with painful PTTs (i.e. any type of painful rotator cuff disorder). Therefore, the prediction of the likely outcome of treatment in clinical practice remains difficult. The estimated MID of the WORC can be used to facilitate the interpretation of WORC outcome scores in contexts that are similar to that of my study. The responder analysis provides a means to enhance interpretation by considering the uncertainty in individuals’ true outcome values.

8.3.2 Implications for research

The lack of a clinically usable prognostic model for predicting the outcome of conservative treatment with physiotherapy in adults with painful PTTs (same as with any other type of rotator cuff disorder) highlights the need for further prognostic research. Larger studies are needed. The methodological deficiencies of primary prognostic model studies, as evidenced by the prognostic systematic review, point to the need for further methodologically sound (and adequately reported) studies. Prognosis research is an evolving field, and researchers should be receptive to the developing methods. A growing amount of guidance is available. Researchers need to be conscious of the fact that the provision of a clinically usable prognostic model requires a comprehensive programme of research involving development, (external) validation and assessment of clinical impact (Steyerberg et al. 2013). The methodological advances will hopefully also decrease the heterogeneity of methods which constitutes a considerable barrier to the comparability of studies and their findings, and thereby, to the construction of a homogenous body of evidence.

Besides the need for methodological improvements, the selection of candidate factors for prognostic model studies remains an important research issue. Due to the demonstrated dearth of knowledge on relevant factors within the field of shoulder pain and rotator cuff disorders, the selection of candidate factors for inclusion in a prognostic model study remains difficult. The optimal method for selecting factors is yet unclear. Systematic reviews of prognostic factor studies may facilitate the identification of established or promising factors. The usefulness of systematic reviews, though, depends on the availability of a sufficient number of methodologically sound primary studies. In this context, while I did not evaluate the risk of bias of the prognostic factors studies which I used to inform the selection of factors for my study, prognostic factors studies have been claimed to be commonly affected by methodological deficiencies, which puts them at risk of bias, i.e. of providing invalid results (Riley et al. 2013). Beyond that, the evaluation of risk of bias of prognostic factors studies was until very recently difficult due to a lack of consensus on the assessment of risk of bias in prognostic studies (Hayden et al. 2006, Hayden et al. 2009). Both primary prognostic factor studies and systematic reviews of these studies are likely to benefit from the recently published instruments for the assessment of risk of bias in prognostic factor studies, such as the “Quality In Prognostic Studies” tool (QUIPS) (Hayden et al. 2013). This similarly applies to systematic reviews of prognostic model studies, for which PROBAST (Wolff et al. 2015) is likely to be publicly available soon.

The identification of putative prognostic factors may further be enhanced by consensus studies such as Delphi procedures, i.e. by considering different perspectives such as those of clinical experts, patients and researchers. Consensus is also needed on the definition and standardization of factors and their measurement to lay the foundations for the creation of a sufficiently homogeneous body of evidence.

Further methodologically sound MID studies are needed to provide a range of estimates for use in populations of adults with painful PTTs. Further research is needed to establish the best methods of determining the MID. Researchers need to be aware that reported rates of responders to treatment are likely to vary depending on the responder definition used, and that a responder definition based on the probabilities that an individual's true outcome value is clinically beneficial as defined by the MID and given the typical error of the outcome measure represents a more realistic approach to the interpretation of outcomes.

8.4 REFERENCES

Hayden J. A., Chou R., Hogg-Johnson S., Bombardier C. (2009). Systematic reviews of low back pain prognosis had variable methods and results: guidance for future prognosis reviews. *J Clin Epidemiol*, *62*(8), 781–96.e1. doi:10.1016/j.jclinepi.2008.09.004.

Hayden J. A., Côté P., Bombardier C. (2006). Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med*, *144*(6), 427–37.

Hayden J. A., van der Windt D. A., Cartwright J. L., Côté P., Bombardier C. (2013). Assessing bias in studies of prognostic factors. *Ann Intern Med*, *158*(4), 280–6. doi:10.7326/0003-4819-158-4-201302190-00009.

Riley R. D., Hayden J. A., Steyerberg E. W., Moons K. G. M., Abrams K., Kyzas P. A., Malats N., Briggs A., Schroter S., Altman D. G., Hemingway H., PROGRESS Group. (2013). Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*, 10(2), e1001380. doi:10.1371/journal.pmed.1001380.

Schellingerhout J. M., Verhagen A. P., Thomas S., Koes B. W. (2008). Lack of uniformity in diagnostic labeling of shoulder pain: time for a different approach. *Man Ther*, 13(6), 478–83. doi:10.1016/j.math.2008.04.005.

Steyerberg E. W., Moons K. G. M., van der Windt D. A., Hayden J. A., Perel P., Schroter S., Riley R. D., Hemingway H., Altman D. G. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med*, 10(2), e1001381. doi:10.1371/journal.pmed.1001381.

Wolff R., Whiting P., Moons K. (2015). *Systematic reviews of prognostic studies 1: assessing risk of bias in studies of prediction models using the PROBAST tool.* Workshop at 23rd Cochrane Colloquium, Vienna, 3-7 Oct 2015. Available at: http://2015.colloquium.cochrane.org/workshops/systematic-reviews-prognostic-studies-1-assessing-risk-bias-studies-prediction-models [Last accessed 03 June 2016].

PART FOUR:

APPENDICES

Appendix 3.1
Braun et al. 2015

Due to copyright issues, this thesis version does not include appendices 3.1 and 4.1.

Appendix 3.2
Review protocol

Braun et al., Teesside University, 02 April 2014

Protocol for a prognostic systematic review

Authors/contributors
Braun C.* **, Hanchard N.C.A.*, Handoll H.H.G.*, Batterham A.M.*, Betthaeuser A.***; *Health and Social Care Institute, Teesside University, Middlesbrough, UK; **hochschule 21, Buxtehude, Germany, ***Schulterzentrum, Hamburg, D.

Title
Prognostic models for the outcome of conservative treatment including physiotherapy in adults with rotator cuff disorders: a systematic review

1. Background and rationale

1.1. Description of disease

Shoulder complaints rank among the commonest musculoskeletal disorders and are frequently seen in medical and physiotherapy practice (Feleus et al. 2008; Karels et al. 2006; Kooijman et al. 2013). Disorders related to the structures within the subacromial space, i.e. the rotator cuff and the subacromial-subdeltoid bursae, have consistently been reported to form the largest subgroup (44% to 89%) of shoulder disorders in clinical practice (Kooijman et al. 2013; van der Windt et al. 1995; van der Windt et al. 1996; Virta et al. 2012). The rotator cuff, a deep cuff of four tendons around the shoulder, is often involved in a progressive degenerative continuum (Cook & Purdam 2009; Lewis 2010). Widely used diagnostic labels for disorders related to the rotator cuff include “subacromial pain”, “shoulder impingement” or “rotator cuff disease”. These encompass a wide range of structural pathologies from rotator cuff tendinopathy to partial or full-thickness rotator cuff tears. The overall prevalence of rotator cuff tears in the general population has been reported as > 40% (Reilly et al. 2006), and is strongly associated with increasing age (Beaudreuil et al. 2007). Rotator cuff-related disorders can significantly impair shoulder function and health-related quality of life (Ryliskis et al. 2009), and can lead to prolonged sick leave (Virta et al. 2012).

1.2. Diagnosis

Clinical signs and symptoms associated with rotator cuff disorders include pain in the shoulder with movements of the arm or at rest, pain-related impairment of shoulder

function and, in more advanced cases (specifically in the presence of a rotator cuff tear), abnormalities on tests of rotator cuff function and integrity (Hanchard et al. 2013). Verification of the presence of a rotator cuff tear requires diagnostic imaging (e.g. ultrasonography (US), magnetic resonance imaging (MRI)) (Lenza et al. 2013; Smith et al. 2011).

1.3. Treatment

Current international guidelines for the management of rotator cuff disorders recommend conservative treatment as the first-line treatment, with surgery reserved for non-responders (Beaudreuil et al. 2010; Haute Autorité de Santé 2005; Robb et al. 2009; American Academy of Orthopaedic Surgeons 2011), or for patients with full-thickness rotator cuff tears, in whom early consideration of surgery is commonly recommended.

A variety of conservative and surgical options is available (Braun et al. 2013; Gebremariam et al. 2011; Littlewood et al. 2011; Hanratty et al. 2012; Seida et al. 2010). Conservative treatment usually consists of medical care (e.g. advice, oral medication, corticosteroid injections) as well as physiotherapy (e.g. exercises, manual therapy). Surgical treatment options include a variety of arthroscopic, mini-open or open approaches to subacromial decompression, debridement or repair. Surgical intervention is a more invasive approach, and not without risk such as infection and iatrogenic injury. Even though the overall rate of complications of rotator cuff surgery appears to be low, shoulder arthroscopic procedures also carry a small but real potential for life-threatening complications (Marecek & Saltzman 2010; Randelli et al. 2012).

1.4. What is known

Despite the available range of treatment options the published literature on the management of rotator cuff disorders reflects a considerable uncertainty about the specific indications for any treatment, whether conservative or surgical (Beaudreuil et al. 2010; Haute Autorité de Santé 2005; Robb et al. 2009; American Academy of Orthopaedic Surgeons 2011). A growing number of studies provide evidence that both conservative and surgical approaches can lead to successful outcomes in patients with any rotator cuff disorder, but there is a paucity of direct comparisons. As yet, only a small number of studies have investigated the comparative effectiveness

of conservative versus surgical treatment (Braun et al. 2013; Dorrestijn et al. 2009; Gebremariam et al. 2011; Tashjian 2013; Kukkonen et al. 2014). Overall, these studies provide no evidence for any clinically relevant differences in outcomes between conservative and surgical treatment. Nonetheless, there is a well-documented increase in the rates of surgical interventions for rotator cuff disorders in many countries (Colvin et al. 2012; Svendsen et al. 2012; Yu et al. 2010).

1.5. Prognostic research and rationale for the planned review

Unnecessary surgery is obviously undesirable, but so is ineffective conservative treatment. It would consequently benefit both patients and health care providers if likely responders and, by corollary, non-responders to conservative interventions, could be identified at the commencement of their care pathway. Early identification would help preserve patients from unnecessary or prolonged suffering, reduce uncertainty and anxiety and limit exposure to the risk of surgery. Also, optimization of treatment selection would save time and effort of both patients and health professionals, and would promote the optimal distribution of available resources.

The importance of predicting which patients will respond to particular treatments has recently been recognised, prompting an increased interest in musculoskeletal prognostic research (Stanton et al. 2010). There has been a corresponding and timely development in related prognostic research methodology, including prognostic models to guide treatment decisions (Moons et al. 2009; PROGRESS 2013; Cochrane Prognosis Methods Group 2013). Prognostic models, which are derived from combinations of multiple prognostic factors (a prognostic factor being “any measure that, among people with a given health condition, is associated with a subsequent clinical outcome“, (PROGRESS 2013)), aim to predict the risk of future clinical outcomes in patients or healthy people. Research on individual prognostic models encompasses three phases: 1) the development of a model (including its internal validation); 2) the external validation of the model; and 3) the investigation of the clinical impact of the model (Steyerberg et al. 2013).

To our knowledge, this is the first systematic review to synthesize the available evidence on prognostic models for the outcome of physiotherapy in adults with rotator cuff disorders.

2. Objectives

This review aims to identify, evaluate and synthesize the available research on prognostic models for the outcome conservative treatment including physiotherapy in adults with shoulder pain and coexistent rotator cuff disorders.

By this, we aim to provide a concise resource to facilitate clinical decision-making but also to identify any research gaps.

3. Methods:

3.1. General

This protocol was developed in the light of current methodological recommendations by the Cochrane Prognosis Methods Group (Cochrane Prognosis Methods Group 2013) and the PROGgnosis RESearch Strategy (PROGRESS) Partnership (PROGRESS 2013) as well as related relevant publications (e.g. (Altman 2001; Altman 2009; Bouwmeester et al. 2012; Geersing et al. 2012; Hayden et al. 2006; Hayden et al. 2009; Hayden et al. 2013; Hemingway et al. 2013; Huguet et al. 2013; Steyerberg et al. 2013). This review protocol will be registered in the international Prospective Register of Systematic Reviews, PROSPERO (PROSPERO 2013).

3.2. Criteria for selecting studies for this review

3.2.1. Types of studies

This review will include primary studies specifically designed to explore prognostic models for the outcome of conservative treatment including physiotherapy in adults with rotator cuff disorders. According to the PROGRESS partnership (PROGRESS 2013), prognostic models „utilise multiple prognostic factors in combination to predict the risk of future clinical outcomes in individual patients“. Inclusion will encompass all phases of prognostic model research (PROGRESS 2013; Steyerberg et al. 2013), i.e. any studies designed to develop, validate or assess the clinical impact of prognostic models. We will consider any longitudinal research designs, but inclusion will be restricted to prospective studies, including randomised controlled trials (RCTs). Inclusion will also be restricted to study reports published in English, as our resources do not allow for inclusion of reports in other languages (but see 3.3.2. for approach to searches regardless of language, and documentation of findings).

3.2.2. Types of participants

We will include studies on populations of adult patients (age ≥ 18 years) who have a diagnosis of a non-traumatic rotator cuff disorder. We define a rotator cuff disorder as shoulder pain that is related to an impairment or dysfunction of the rotator cuff tendons (supraspinatus, infraspinatus, teres minor or subscapularis). Diagnoses of tendinitis, tendonitis, tendinopathy or tear, as applied to these tendons, are our explicit focus. We will also include studies whose inclusion criteria are symptoms or mechanisms consistent with rotator cuff disorders e.g. subacromial pain, or subacromial or shoulder impingement. We will not actively seek studies concentrating on subacromial–subdeltoid bursitis *per se* although, due to its intimate relationship to the rotator cuff, incidental involvement of this bursa may well occur in our population of interest.

We will include research on the full spectrum of rotator cuff disorders as defined above, and as diagnosed by clinical examination and/or diagnostic imaging (e.g. ultrasonography (US), magnetic resonance imaging (MRI) or magnetic resonance arthrography (MRA)). We define a non-traumatic disorder as a disorder that is considered unrelated to a substantial trauma involving the shoulder (e.g. shoulder dislocation). No restrictions will be made on the duration or severity of symptoms at the time of baseline assessment. We will include studies regardless of the setting of care (e.g. primary or secondary care, inpatient or outpatient settings).

We will exclude studies focusing on patients who are pain-free or have trauma-related conditions. We will also exclude studies on disorders of the long head of biceps or calcific tendinitis. Ideally, included studies will specifically exclude other potential sources of shoulder pain (e.g. frozen (contracted) shoulder, glenoid labrum pathologies, previous substantial shoulder trauma, previous surgery at the affected shoulder, neck disorders, multisite musculoskeletal pain, relevant systemic diseases and disorders, or neurologic disorders). Studies in which 85% or more of participants satisfy our criteria will also be included. We anticipate that in some studies there will be insufficient characterisation of participants (e.g. other potential causes of shoulder pain might not be considered). The impact of including these studies will be evaluated by sensitivity analyses.

3.2.3. Types of interventions

Inclusion will be restricted to studies evaluating conservative treatment including physiotherapy with or without adjunctive medical treatment (e.g. advice, oral analgesics, steroid injections) as part of a non-surgical care pathway. We define physiotherapy as any type of exercises and/or manual mobilisation as commonly supplied by physiotherapists. Additional physical therapy modalities (e.g. electrotherapy, thermotherapy) may be an additional part of the physiotherapy treatment, but only if supplied as a supplement to exercises and/or manual mobilisation. Adjunctive treatments that are usually not considered core elements of physiotherapy practice (e.g. acupuncture or osteopathic musculoskeletal interventions) are permissible. Studies on non-surgical interventions that do not include physiotherapy (e.g. corticosteroid injections alone) will be excluded. No restrictions will be made on the duration or frequency of the physiotherapy. Studies with two or more groups, in which any other non-surgical, surgical or no treatment is compared with non-surgical treatment including physiotherapy, will only be considered if there is separate prognostic modelling for the latter.

.

3.2.4. Types of prognostic factors

We will include studies that model potential prognostic factors elicited at the baseline assessment. We anticipate that these factors will typically include demographic (e.g. age or shoulder function), clinical (e.g. measures of symptom severity) or diagnostic imaging (e.g. type of structural rotator cuff defect) characteristics. Studies modelling potential prognostic factors that were not elicited from the baseline assessment will be excluded.

3.2.5. Types of outcomes

Primary outcomes addressed by this review will be

- Pain
- Shoulder disability (as measured by patient-reported outcome measures (PROMs, e.g. Shoulder Pain and Disability Index or Western Ontario Rotator Cuff Index))
- Adverse events

Secondary outcomes will be

- Health-related quality of life (HrQoL)

- Sick leave (i.e. time off work due to the shoulder problem)
- Global perceived change (GPC)
- Structural progression from no tear or partial tears to full-thickness tears (as determined by US, MRI or MRA)

Any eligible study must present a prognostic model in relation to at least one of these outcomes. Ideally, study authors should provide information on the psychometric properties of all measurements used.

3.2.6. Types of analysis

For inclusion, studies must include or evaluate a prognostic model of multiple prognostic factors, but no restriction will be placed on the phase (modelling, validation or evaluation of impact) or the type of analysis. Details will be documented.

3.3. Search methods for identification of studies

3.3.1. Search strategy

Acknowledging the known difficulties with the retrieval of prognostic model research in electronic databases (Geersing et al. 2012; Walker-Dilks et al. 2008), we have given careful attention to the development of a sensitive search strategy with which we expect to detect most of the available prognostic model studies that are relevant to our review question. The final search strategy was informed by a preliminary strategy that we had developed in 2011 preparatory to a prognostic study in the field. We have revised and updated this strategy for the purpose of the present review. We tested multiple combinations of search terms for the population, interventions, comparisons, prognostic factors and outcomes, but decided on a broad strategy including only search terms relating to or describing the population and interventions. The final selection of search terms was informed by findings from test searches, as well as by the experience of previous searches for two intervention systematic reviews in the field (Braun & Hanchard 2010; Braun et al. 2013). We further tested various available search filters for the identification of prognostic research (e.g. (Altman 2001; Geersing et al. 2012; Haynes et al. 2005; Wilczynski & Haynes 2004; Wilczynski & Haynes 2005; Walker-Dilks et al. 2008). For the Medline search we will use a recently developed and validated search string for clinical prediction models studies by Geersing et al. (2013, with a small supplement) but, due to concerns about the currency and performance of some of the available filters, we decided to

search without such filters in the other databases. See Appendix for the Medline (EBSCO format) strategy.

We will search the following electronic databases: Medline (PubMed), Embase, Cochrane CENTRAL, Cinahl and PEDro. We will further search the WHO International Clinical Trials Registry Platform (ICTRP). We are aware of the relevance of further sources such as citation indices, hand searches and grey literature, but our resources do not allow for their consideration for this review. The databases' "related articles" functions will be applied to all relevant findings. We will further search the bibliographies of relevant articles and existing systematic reviews. Even though inclusion will be restricted to publications in the English language (see also 3.2.1.), we will search regardless of language and document any findings in other languages. We will incorporate the findings from our previous (2011) searches.

3.3.2. Search and selection process

The process of study selection will follow current methodological guidelines (Higgins & Green 2011) and will be documented by a PRISMA flow diagram (Moher et al. 2010). One author (CB) will conduct the searches and two authors (CB, NH) will independently perform study selection. The findings will then be compared. In case of disagreement, consensus will be sought through discussion, or through involvement of a third independent researcher. Reasons will be given for all final exclusions.

3.4. Data collection and analyses

3.4.1. Data collection

Two researchers (CB, NH) will independently collect data using a piloted data extraction form. Data extraction will include the following key items:

- Publication details: e.g. authors, year of publication
- Study design: type of study and phase of research (i.e. model development, external validation or investigations of clinical impact)
- Study characteristics: e.g. year and place of study, setting (primary or secondary care, inpatient or outpatient)
- Sample size and sample size justification
- Population characteristics: type of rotator cuff disorder, inclusion/exclusion criteria, sample characteristics (e.g. age, sex, employment)
- Intervention characteristics: e.g. content, duration, dosage, compliance

- Duration of follow-up, losses to follow-up
- Outcomes: primary and secondary outcomes, outcome measurements
- Prognostic factors: numbers, definitions and details of factors and factor measurements
- Analyses: details of methods of multivariable prognostic modelling (including any adjustments and methods of dealing with missing data)
- Results for each outcome: final statistical model(s), predictive performance/accuracy statistics with their confidence intervals. All prognostic factors considered in the analyses will be presented regardless of "statistical significance".

Where necessary, we will attempt to contact study authors for unreported study details and data. We will not impute missing data.

3.4.2. Assessment of study quality and risk of bias

We will use the Prediction Model Studies Risk of Bias Tool (PROBAST 2013) for the assessment of risk of bias in the included prognostic model studies. We have been in contact with Robert Wolff (a member of the PROBAST steering group) on the use of this new instrument (Wolff 2013). Publication of the instrument is planned for the first half of 2014. The PROBAST tool is being developed to assess risk of bias in prediction model studies and includes the following five key domains: 1) participant selection, 2) outcome, 3) predictors, 4) sample size and flow, 5) analysis. Depending on the stage of development of PROBAST at the time when we will assess the studies for our review, we will either use the current pre-final version, or the final version. We will conduct supplementary risk of bias assessments for other study types, using appropriate instruments (e.g. the Cochrane Collaboration's risk of bias tool (Higgins & Green 2011) for RCTs investigating the clinical impact of a prognostic model).

Risk of bias will be assessed independently by two researchers (CB, NH). In case of disagreement, consensus will be sought through discussion, or through involvement of a third independent researcher.

3.4.4. Analysis and synthesis

We will provide a narrative summary of the findings of all included studies and for each outcome. In particular, this will include the presentation of all prognostic factors included in the final prognostic model, specification of the prognostic model, its actual outcome as well as its prognostic accuracy. We are planning to categorize the results by the type of cuff disorder (non-tear populations (rotator cuff tendinitis/"subacromial impingement"/"subacromial pain"), PTT, FTT, or mixed populations), and by type of treatment (physiotherapy alone versus physiotherapy plus adjunctive medical treatment). Should we decide on any other categorisation, we will note this and provide a rationale. We will synthesise the results in consideration of the phase of research, i.e. according to if (and how) the presented models have been externally validated or tested for clinical impact.

If a sufficient number of good quality studies on the same prognostic model are available, we will summarise the performance of the model through a meta-analysis. If appropriate, exploratory meta-regression will be performed to assess sub-group and continuous covariate effects (such as type of disorder, sex or treatment parameters).

Acknowledgements

We would like to express our special thanks to Robert Wolff, Kleijnen Systematic Reviews Ltd., York, UK, for his valuable feedback on the review protocol, and to Iain Baird and Julie Hogg, Library and Information Services, Teesside University, UK, for their very helpful support in developing the search strategy.

Declarations of interest

The authors are currently involved in an ongoing (non-funded) study aimed at developing a prognostic model for the outcome of conservative treatment including physiotherapy in adults with partial-thickness rotator cuff tears.

4. References

Altman, D.G. (2009) Prognostic models: a methodological framework and review of models for breast cancer. *Cancer Investigation*, 27(3), pp.235–43.

Altman, D.G. (2001) Systematic reviews of evaluations of prognostic variables. *BMJ (Clinical Research ed.)*, 323(7306), pp.224–8.

American Academy of Orthopaedic Surgeons (2011) AAOS clinical practice guideline: optimizing the management of rotator cuff problems. *The Journal of the American Academy of Orthopaedic Surgeons*, 19(6), pp.380–3.

Beaudreuil, J., Bardin, T., Orcel, P. & Goutallier, D. (2007) Natural history or outcome with conservative treatment of degenerative rotator cuff tears. *Joint, Bone, Spine*, 74(6), pp.527–9.

Beaudreuil, J., Dhénain, M., Coudane, H. & Mlika-Cabanne, N. (2010) Clinical practice guidelines for the surgical management of rotator cuff tears in adults. *Orthopaedics & Traumatology: Surgery & Research*, 96(2), pp.175–179.

Bouwmeester, W., Zuithoff, N.P.A., Mallett, S., Geerlings, M.I., Vergouwe, Y., Steyerberg, E.W., Altman, D.G. & Moons, K.G.M. (2012) Reporting and methods in clinical prediction research: a systematic review. *PLoS Medicine*, 9(5), pp.1–12.

Braun, C., Bularczyk, M., Heintsch, J. & Hanchard, N.C.A. (2013) Manual therapy and exercises for shoulder impingement revisited. *Physical Therapy Reviews*, 18(4), pp.263–284.

Braun, C. & Hanchard, N.C.A. (2010) Manual therapy and exercise for impingement-related shoulder pain. *Physical Therapy Reviews*, 15(2), pp.62–83.

Cochrane Prognosis Methods Group (2013) Cochrane Prognosis Methods Group. Online at: http://prognosismethods.cochrane.org/ [last accessed March 22, 2014].

Colvin, A.C., Egorova, N., Harrison, A.K., Moskowitz, A. & Flatow, E.L. (2012) National trends in rotator cuff repair. *The Journal of Bone and Joint Surgery. American volume*, 94(3), pp.227–33.

Cook, J.L. & Purdam, C.R. (2009) Is tendon pathology a continuum? A pathology model to explain the clinical presentation of load-induced tendinopathy. *British Journal of Sports Medicine*, 43(6), pp.409–16.

Dorrestijn, O., Stevens, M., Winters, J.C., van der Meer, K. & Diercks, R.L. (2009) Conservative or surgical treatment for subacromial impingement syndrome? A systematic review. *Journal of Shoulder and Elbow Surgery*, 18(4), pp.652–60.

Feleus, A., Bierma-Zeinstra, S.M.A., Miedema, H.S., Bernsen, R.M.D., Verhaar, J.A.N. & Koes, B.W. (2008) Incidence of non-traumatic complaints of arm, neck and shoulder in general practice. *Manual Therapy*, 13(5), pp.426–33.

Gebremariam, L., Hay, E.M., Koes, B.W. & Huisstede, B.M. (2011) Effectiveness of surgical and postsurgical interventions for the subacromial impingement syndrome: a systematic review. *Archives of Physical Medicine and Rehabilitation*, 92(11), pp.1900–13.

Geersing, G.-J., Bouwmeester, W., Zuithoff, P., Spijker, R., Leeflang, M. & Moons, K. (2012) Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PloS One*, 7(2), e32844.

Hanchard, N.C., Lenza, M., Handoll, H.H. & Takwoingi, Y. (2013) Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement. *Cochrane Database of Systematic Reviews*, 4, CD007427.

Hanratty, C.E., McVeigh, J.G., Kerr, D.P., Basford, J.R., Finch, M.B., Pendleton, A. & Sim, J. (2012) The Effectiveness of Physiotherapy Exercises in Subacromial Impingement Syndrome: A Systematic Review and Meta-Analysis. *Seminars in Arthritis and Rheumatism*, 42(3), pp.297–316.

Haute Autorité de Santé (2005) Clinical practice guidelines Management of chronic painful shoulder without instability in adults. Available at: http://www.has-sante.fr/portail/upload/docs/application/pdf/painful_shoulder_guidelines.pdf [last accessed March 22, 2014].

Hayden, J.A., Chou, R., Hogg-Johnson, S. & Bombardier, C. (2009) Systematic reviews of low back pain prognosis had variable methods and results: guidance for future prognosis reviews. *Journal of Clinical Epidemiology*, 62(8), pp.781–796.e1.

Hayden, J.A., Côté, P. & Bombardier, C. (2006) Evaluation of the quality of prognosis studies in systematic reviews. *Annals of Internal Medicine*, 144(6), pp.427–37.

Hayden, J.A., van der Windt, D.A., Cartwright, J.L., Côté, P. & Bombardier, C. (2013) Assessing bias in studies of prognostic factors. *Annals of Internal Medicine*, 158(4), pp.280–6.

Haynes, R.B., Kastner, M. & Wilczynski, N.L. (2005) Developing optimal search strategies for detecting clinically sound and relevant causation studies in EMBASE. *BMC Medical Informatics and Decision Making*, 5, 8.

Hemingway, H., Croft, P., Perel, P., Hayden, J.A., Abrams, K., Timmis, A., Briggs, A., Udumyan, R., Moons, K.G.M., Steyerberg, E.W., Roberts, I., Schroter, S., Altman, D.G. & Riley, R.D. (2013) Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ*, 346, e5595.

Higgins, J.P.T.; Green, S. (2011) *Cochrane Handbook for Systematic Reviews of Interventions*. Online at: http://www.cochrane.org/cochrane-interventions-handbook [last accessed March 22, 2014].

Huguet, A., Hayden, J. a, Stinson, J., McGrath, P.J., Chambers, C.T., Tougas, M.E. & Wozney, L. (2013) Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. *Systematic Reviews*, 2(1), p.71.

Karels, C.H., Polling, W., Bierma-Zeinstra, S.M.A., Burdorf, A., Verhagen, A.P. & Koes, B.W. (2006) Treatment of arm, neck, and/or shoulder complaints in physical therapy practice. *Spine*, 31(17), pp.E584–9.

Kooijman, M., Swinkels, I., van Dijk, C., de Bakker, D. & Veenhof, C. (2013) Patients with shoulder syndromes in general and physiotherapy practice: an observational study. *BMC Musculoskeletal Disorders*, 14(1), 128.

Kukkonen, J., Joukainen, A., Lehtinen, J., Mattila, K.T., Tuominen, E.K.J., Kauko, T. & Aärimaa, V. (2014) Treatment of non-traumatic rotator cuff tears: A randomised controlled trial with one-year clinical results. *The Bone & Joint Journal*, 96(1), pp.75–81.

Lenza, M., Buchbinder, R., Takwoingi, Y., Johnston, R.V., Hanchard, N.C.A. & Faloppa, F. (2013) Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered. *The Cochrane Database of Systematic Reviews*, 9, CD009020.

Lewis, J.S. (2010) Rotator cuff tendinopathy: a model for the continuum of pathology and related management. *British Journal of Sports Medicine*, 44(13), pp.918–23.

Littlewood, C., Ashton, J., Chance-Larsen, K., May, S. & Sturrock, B. (2011) Exercise for rotator cuff tendinopathy: a systematic review. *Physiotherapy*, 98(2), pp.101–109.

Marecek, G.S. & Saltzman, M.D. (2010) Complications in shoulder arthroscopy. *Orthopedics*, 33(7), pp.492–7.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D.G. (2010) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International Journal of Surgery*, 8(5), pp.336–41.

Moons, K.G.M., Royston, P., Vergouwe, Y., Grobbee, D.E. & Altman, D.G. (2009) Prognosis and prognostic research: what, why, and how? *BMJ*, 338, b375.

PROABAST (2013) Prediction Risk of Bias Tool. Online at: www.systematic-reviews.com/probast [last accessed March 22, 2014].

PROGRESS (2013) PROGnosis RESearch Strategy Partnership. Online at: http://www.progress-partnership.org/ [last accessed March 22, 2014].

PROSPERO (2013) International Prospective Register of Systematic Reviews. Online at http://www.crd.york.ac.uk/PROSPERO/ [last accessed March 22, 2014].

Randelli, P., Spennacchio, P., Ragone, V., Arrigoni, P., Casella, A. & Cabitza, P. (2012) Complications associated with arthroscopic rotator cuff repair: a literature review. *Musculoskeletal Surgery*, 96(1), pp.9–16.

Reilly, P., Macleod, I., Macfarlane, R., Windley, J. & Emery, R.J.H. (2006) Dead men and radiologists don't lie: a review of cadaveric and radiological studies of rotator cuff tear prevalence. *Annals of the Royal College of Surgeons of England*, 88(2), pp.116–21.

Robb, G., Arroll, B., Reid, D. & Goodyear-Smith, F. (2009) Summary of an evidence-based guideline on soft tissue shoulder injuries and related disorders – Part 1: Assessment. *Journal of Primary Health Care*, 1(1), pp.36–41.

Ryliskis, S., Brophy, R.H., Kocius, M. & Marx, R.G. (2009) Shoulder activity level in the preoperative assessment of patients with rotator cuff tears. *Knee Surgery, Sports Traumatology, Arthroscopy*, 17(12), pp.1522–8.

Seida, J.C., LeBlanc, C., Schouten, J.R., Mousavi, S.S., Hartling, L., Vandermeer, B., Tjosvold, L. & Sheps, D.M. (2010) Systematic review: nonoperative and operative treatments for rotator cuff tears. *Annals of Internal Medicine*, 153(4), pp.246–55.

Smith, T.O., Back, T., Toms, A.P. & Hing, C.B. (2011) Diagnostic accuracy of ultrasound for rotator cuff tears in adults: A systematic review and meta-analysis. *Clinical Radiology*, 66(11), pp.1036–48.

Stanton, T.R., Hancock, M.J., Maher, C.G. & Koes, B.W. (2010) Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. *Physical Therapy*, 90(6), pp.843–54.

Steyerberg, E.W., Moons, K.G.M., van der Windt, D.A., Hayden, J.A., Perel, P., Schroter, S., Riley, R.D., Hemingway, H. & Altman, D.G. (2013) Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PloS Medicine*, 10(2), e1001381.

Svendsen, S.W., Frost, P. & Jensen, L.D. (2012) Time trends in surgery for non-traumatic shoulder disorders and postoperative risk of permanent work disability: a nationwide cohort study. *Scandinavian Journal of Rheumatology*, 41(1), pp.59–65.

Tashjian, R.Z. (2013) Is there Evidence in Favor of Surgical Interventions for the Subacromial Impingement Syndrome? *Clinical Journal of Sport Medicine*, 23(5), pp.406–7.

Virta, L., Joranger, P., Brox, J.I. & Eriksson, R. (2012) Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. *BMC Musculoskeletal Disorders*, 13(1), 17.

Walker-Dilks, C., Wilczynski, N.L. & Haynes, R.B. (2008) Cumulative Index to Nursing and Allied Health Literature search strategies for identifying methodologically sound causation and prognosis studies. *Applied Nursing Research*, 21(2), pp.98-103.

Wilczynski, N.L. & Haynes, R.B. (2004) Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. *BMC Medicine*, 2, 23.

Wilczynski, N.L. & Haynes, R.B. (2005) Optimal search strategies for detecting clinically sound prognostic studies in EMBASE: an analytic survey. *Journal of the American Medical Informatics Association*, 12(4), pp.481–5.

Van der Windt, D. a, Koes, B.W., Boeke, a J., Devillé, W., De Jong, B. a & Bouter, L.M. (1996) Shoulder disorders in general practice: prognostic indicators of outcome. *The British Journal of General Practice*, 46(410), pp.519–23.

Van der Windt, D. a, Koes, B.W., de Jong, B. a & Bouter, L.M. (1995) Shoulder disorders in general practice: incidence, patient characteristics, and management. *Annals of the Rheumatic Diseases*, 54(12), pp.959–64.

Wolff, R. (2013) Prediction study risk of bias assessment tool (PROBAST). Online at: http://2013.colloquium.cochrane.org/abstracts/prediction-study-risk-bias-assessment-tool-probast [last accessed March 22, 2014].

Yu, E., Cil, A., Harmsen, W.S., Schleck, C., Sperling, J.W. & Cofield, R.H. (2010) Arthroscopy and the dramatic increase in frequency of anterior acromioplasty from 1980 to 2005: an epidemiologic study. *Arthroscopy*, 26(9 Suppl), pp.S142–7.

Braun et al., Teesside University, 02 April 2014

Appendix: Example search strategy: Medline (compatible with EBSCO)

Prognosis research filter: Geersing et al. (2012) (clinical prediction models studies, Ingui filter OR update (S3 OR S4))

S1	((MH “Shoulder” OR MH “Shoulder Pain” OR shoulder) AND (MH Tendinopathy OR (“soft tissue” OR tendon* OR tendin* OR imping* OR rotator OR cuff).ti,ab)) OR (supraspinatus OR infraspinatus OR “teres minor” OR subscapularis OR „rotator cuff“ OR subacromial*).ti,ab OR MH “Shoulder Impingement Syndrome” OR MH “Rotator Cuff”
S2	MH “Physical Therapy Modalities+” OR MH “Rehabilitation+” OR ("physical therap*" or physiotherap* OR exercis* OR "manual therap*" OR "manipulative therap*" OR mobilis* or rehab* OR conservative* OR non-operat* OR nonoperat* OR non-surg* OR nonsurg*).ti,ab
S3	validat* OR TI predict*.ti OR rule* OR (predict* AND (outcome* OR risk* OR model*)) OR ((history OR variable* OR criteria OR scor* OR characteristic* OR finding* OR factor*) AND (predict* OR model* OR decision* OR identif* OR prognos*)) OR (decision* AND (model* OR clinical* OR MH “Logistic Models”)) OR (prognostic AND (history OR variable* OR criteria OR scor* OR characteristic* OR finding* OR factor* OR model*))
S4	stratification OR MH "ROC Curve" OR discrimination OR discriminate OR c-statistic OR "c statistic" OR area under the curve OR AUC OR calibration OR indices OR algorithm OR multivariable
S5	prognos*.ti,ab
S6	S1 AND S2 AND (S3 OR S4 OR S5)

Appendix 3.3
Filter for screening of titles and abstracts

In the title or abstract:

3) Key terms in the free text, and the context in which they are used, suggest that prognostic modelling may have been conducted.

   Examples of key terms include “prognosis”, “predictor”, “follow-up” and “correlation” among numerous others.

4) It appears that the study may have been a longitudinal cohort study, an RCT or a quasi-RCT.

5) It appears that the study may have evaluated impingement and or rotator cuff disease.

6) It appears that the study evaluated a conservative intervention including physiotherapy.

7) It appears that the study was, or may have been, prospective.

Tend towards over-inclusivity in cases of uncertainty.

Rate as:

“yes” (includes “unsure”): retain record and obtain full text; or

“no”: delete record.

Appendix 3.4

Eligibility form for screening of full texts

Data collection form – 2nd screening/ ELIGIBILITY 10 July_2014

Summary judgment (see p.3): INCLUDE ☐ EXCLUDE ☐ DISCUSS ☐

Review title	*Prognostic models for the outcome of conservative treatment including physiotherapy in adults with rotator cuff disorders: a systematic review*
Study ID *(surname of first author and year first full report of study was published e.g. Smith 2001)*	
Reference of (primary) report	*See Mendeley folder*
Report ID of other reports of this study including errata or retractions	
Notes	

General Information

Date of assessment	
Name of person extracting data	Cordula ☐ Nigel ☐
Study author contact details	
Publication type	Full text ☐
Notes:	

Adopted from:
http://www.cochrane.org/sites/default/files/uploads/forums/u389/ERC%20data%20collection%20form%20for%20intervention%20reviews%20for%20RCTs%20only.doc (26 March 2014)

Study eligibility

Study Characteristics	Eligibility criteria	Eligibility criteria met? Yes	No	Unclear	Location in text or source
Type of study	(Longitudinal) cohort study	☐	☐	☐	
	RCT or quasi-RCT	☐	☐	☐	
	Prospective design	☐	☐	☐	
Participants	Inclusion criteria. Please circle if applicable to > 85%: • Adults age =/> 18 • Shoulder pain and • Diagnosis of a rotator cuff disorder (tendinitis, tendonitis, tendinopathy, subacromial pain, subacromial or shoulder impingement, rotator cuff tears) • Non-traumatic condition Exclusion criteria. Please circle if applicable to > 85%: • Pain-free population • Traumatic condition • Long head of biceps disorder(s) • Calcific tendinitis • Other sources of shoulder pain	☐	☐	☐	
Types of intervention	Conservative treatment including physiotherapy?	☐	☐	☐	
Prognostic factors	Set of potential prognostic factors (min. 2) been investigated within this study?	☐	☐	☐	
Types of Prognostic factors	All factors elicited from baseline assessment?	☐	☐	☐	

Study characteristics	Eligibility criteria	Eligibility criteria met? Yes	No	Unclear	Location in text or source
Outcome measures	At least one of the following included (please circle)? Primary: • Pain • Disability • Adverse events Secondary: • Health-related Quality of Life • Sick leave • Global perceived change • Structural progression (FTT) • Need for surgery	☐	☐	☐	
Analysis	Prognostic modelling of multiple prognostic factors?	☐	☐	☐	
In RCTs/studies with 2 or more groups	Is there separate prognostic modelling for any group(s) receiving conservative treatment with physiotherapy?	☐	☐	☐	(cross out if not applicable)
INCLUDE ☐	EXCLUDE ☐	DISCUSS ☐			
Reason(s) for exclusion or discussion					
Notes:					

?

Appendix 3.5

Data extraction form 1 - model development studies

Data extraction form 1: Prognostic Model *DEVELOPMENT* Studies (with or without internal validation), 10 Dec 2014

Helpful/accompanying background documents:

- Altman, D.G., McShane, L.M., Sauerbrei, W. & Taube, S.E. (2012) Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS Medicine*, 9(5), p.e1001216.
- Mallett, S., Royston, P., Dutton, S., Waters, R. & Altman, D.G. (2010) Reporting methods in studies developing prognostic models in cancer: a review. *BMC Medicine*, 8, p.20.
- Mallett, S., Royston, P., Waters, R., Dutton, S. & Altman, D.G. (2010) Reporting performance of prognostic models in cancer: a review. *BMC Medicine*, 8, p.21.
- PROBAST checklist + guidance (Wolff et al. 2014, internal documents)
- REMARK checklist (McShane et al. 2005, see Altman et al. 2012)

Note:
Black = items as taken over/adopted from REMARK (related item nr in brackets at end) and additional items
Blue = items as taken over from PROBAST (related phase or domain nr in brackets at end)

Study ID (first author/year):

Data extracted by: CB **Date of data extraction:**

No	Item	Details	Location (p.)	Comments
1	**Introduction**			
1.1	Study objectives (+ any pre-specified hypotheses) (1)			
1.2	Intended use of model (Ph.1)			
1.3	Intended moment in time (Ph.1)			
2	**Materials and methods**			
	Setting and Patients			
2.1	Setting (2) (country, primary/secondary care, place of assessment)			
2.2	Inclusion criteria (2)			
2.3	Exclusion criteria (2)			

2.4	Prior testing (Ph.1)			
2.5	Prior treatment (Ph.1)			
	Treatment			
2.6	Place of treatment (e.g. practice, hospital…)			
2.7	Treatment(s) received – characteristics: Type, content, standardisation, frequency, duration (3); specify if different treatments were received (control groups in RCTs…) Describe … any interventions or treatments during this time (D4)			
2.8	Allocation to *treatment*: randomised (NR) /non-randomised (NR) (3)			
	Study design			
2.9	Type of study (cohort, RCT…)			
2.10	Method of recruitment (e.g. consecutive; including whether any stratification or matching was used) (6)			
2.11	Phase of research: development (DEV) only, development *and* validation DEV/VAL? (Note: Use this form for any internal validation using development sample data; use separate data extraction form (for VAL studies) for any (external) validation using different samples/populations)			
2.12	Time period (dates/overall duration of recruitment) (6)			
2.13	Start point (6) (e.g. initial presentation for shoulder pain to GP…) Were participants enrolled at a similar state of health, or were predictors considered to account for differences? (Ph.3)			
2.14	End point (follow-up): (intended) time of			

	follow-up (6) At what time point was the outcome measured (D3)			
	Outcomes			
2.15	Outcome(s)/definition(s)/measurement(s) (7); specify primary/secondary outcomes where applicable			
2.16	Was data provided on (or a reference case made to) the psychometric properties of (the) *outcome measurement(s)* (i.e. when using e.g. available questionnaires, scales, ROM measurements)?			
2.17	If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome (D3)			
2.18	Was a pre-specified outcome definition used? (D3) (specify for each outcome where applicable)			
2.19	Was the outcome defined and determined in a similar way for all participants? (D3) (specify for each outcome where applicable)			
2.20	Was the outcome determined blind to predictor information? (D3) (specify for each outcome where applicable)			
	Predictors			
2.21	N of predictors initially assessed			
2.22	Predictors initially assessed/definitions/measurements (8) (Provide reference to table (where available) if n of predictors is very high)			
2.23	Rationale for (initial) selection of predictors (e.g. based on systematic literature review, expert consensus, theoretical framework…)			

2.24	Was data provided on (or a reference case made to) the psychometric properties of (the) *predictor measurements* (i.e. when using e.g. available questionnaires, scales, ROM measurements)?			
2.25	Predictors which did not use pre-specified or standard predictor definitions (D2)			
2.26	Were predictors excluded from the outcome definition? (D3)			
2.27	Were predictor measurements blinded to outcome data? (D2)			
2.28	Were predictors defined and assessed in a similar way for all participants in the study? (D2.A.)			
2.29	Sample size rationale/target sample size/target power and effect size (where applicable) (9)			
	Statistical and analysis methods (Model building methods (10))			
2.30	Process and method(s) of *selection* of predictors for inclusion in the multivariable analysis (10) (e.g. (ideally) irrespective of the statistical significance of their association with the outcome?)			
2.31	Handling of predictors in analysis: any coding/categorisation/dichotomisation? (11) If categorisation of predictors was used, was it pre-specified or justified? (D5)			
2.32	Describe how the model was derived... (D5); Statistical modelling methods used *within* multivariable analysis (e.g. backward, stepwise regression…)			
2.33	Process/methods used to develop a prognostic index/score/chart or to create			

	risk groups from the prognostic model (or similar, where presented)				
2.34	Pre-specified adjustments/subgroup analyses				
2.35	Handling of missing data (10) (e.g. imputation method) Describe … methods used for missing data (D4)				
3	**Results**				
	Patient flow (12)				
3.1	N of patients meeting inclusion criteria				
3.2	N of non-inclusions/reasons (12)				
3.3	N of patients selected for inclusion (12)				
3.4	N of dropouts/reasons (12) Describe any participants who were excluded from the analysis (D4)				
3.5	Actual sample size, i.e. *n of patients* included in analysis (for each subgroup where applicable) (12)				
3.6	*N of outcome events* included in analysis (for each outcome and subgroup where applicable)				
3.7	Actual follow-up time (median/mean) Describe the time interval between predictors and outcome measurement… (D4)				
	Sample characteristics (baseline)				
3.8	Characteristics (13)	Age			
		Sex			
		Type/characteristics of disorder (e.g. non-tear, partial or full tears, mixed…)			
		Traumatic origin			
		Duration of symptoms			

		Baseline pain			
		Comorbidities			
		Baseline functional status			
		Dominant arm			
		Employment (e.g. full time, retired…)/work status (sick leave)			
		Other			
		Other			
3.9	Extent of missing values/data (13); Describe missing data on predictors and outcomes… (D4)				
	Analysis and presentation				
3.10	List of predictors included in multivariable analysis				
3.11	Predictors where criteria changed during the study (D2)				
3.12	Details of predictors where data was not provided for all participants (D2)				
3.13	Non constant or subjectively measured predictors where measurement was not blinded to outcome data (D2)				
3.14	Non constant or subjectively measured predictors that were not assessed independently (i.e. without knowledge of each other) (D2)				
3.15	Were all pre-specified outcomes measured and analysed? (D3)				
3.16	Final model and (16) unadjusted effect estimates (e.g. odds ratio, hazard ratio (CI))				
3.17	Model adjustments (where applied) and adjusted effect estimates				
3.18	Prognostic index or prognostic statement				

	Model fitting/performance/accuracy			
3.19	Were the relevant (model) performance measures (e.g. calibration, (re-) classification, net benefit) evaluated? (D5)			
3.20	(Model) *discrimination* (e.g. R^2, D statistic, c-index; (graphically: Kaplan Meier plot)) methods and estimates			
3.21	(Model) *calibration* (e.g. expected/observed events ratio (CI), Hosmer-Lemeshow test) methods and estimates			
3.22	Was *optimism* in the model performance accounted for, e.g. using bootstrapping or shrinkage techniques? (D5)			
3.23	Were *complications* in the data (e.g. competing risks, multiple events per individual) accounted for (appropriately)? (D5)			
3.24	Were *non-linear associations* between predictors and the outcome considered and handled appropriately? (D5)			
3.25	Results of any further investigations, if done (e.g. sensitivity analyses) (18)			
4	*Internal validation (where applicable, same sample data; for any external validation see separate form)*			
4.1	Applied method of validation (e.g. bootstrapping or jack-knive methods (same sample), split sample or cross validation (same population))			
4.2	Validation statistics: estimates, performance (discrimination, calibration)			
4.3	Were any modifications to the model suggested in the light of validation?			
4.4	Was optimism in the model performance			

	accounted for, e.g. using bootstrapping or shrinkage techniques? (D5)			
4.5	Were complications in the data (e.g. competing risks, multiple events per individual) accounted for (appropriately)? (D5)			
5	**Supplementary notes/information**			

Appendix 3.6

Data extraction form 2 - model validation studies

Data extraction form 1: Prognostic Model *VALIDATION* Studies (with or without internal validation), 10 Dec 2014

Helpful/accompanying background documents:

- Altman, D.G., McShane, L.M., Sauerbrei, W. & Taube, S.E. (2012) Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS medicine*, 9(5), p.e1001216.
- Mallett, S., Royston, P., Dutton, S., Waters, R. & Altman, D.G. (2010) Reporting methods in studies developing prognostic models in cancer: a review. *BMC medicine*, 8, p.20.
- Mallett, S., Royston, P., Waters, R., Dutton, S. & Altman, D.G. (2010) Reporting performance of prognostic models in cancer: a review. *BMC medicine*, 8, p.21.
- PROBAST checklist + guidance (Wolff et al. 2014, internal documents)
- REMARK checklist (McShane et al. 2005, see Altman et al. 2012)

Note:
Black = items as taken over/adopted from REMARK (related item nr in brackets at end) and additional items
Blue = items as taken over from PROBAST (related phase or domain nr in brackets at end)

Study ID (first author/year):

Data extracted by: CB **Date of data extraction:**

No	Item	Details	Location (p.)	Comments
1	**Introduction**			
1.1	Study objectives (and any pre-specified hypotheses) (1)			
1.2	Type(s) of validation and approach: *internal* (developmental cohort data, same population) or *external* (new cohort, different population, temporal or geographical…) (Note: Use separate data extraction form (for DEV studies) for any report of a development study that includes any internal validation (same sample data))			
2	**Materials and methods**			

	Setting and Patients			
2.1	Setting (2) (country, primary/secondary care, place of assessment(s))			
2.2	Inclusion criteria (2)			
2.3	Exclusion criteria (2)			
2.4	Prior testing (Ph.1)			
2.5	Prior treatment (Ph.1)			
	Treatment			
2.6	Place of treatment (e.g. practice, hospital…)			
2.7	Treatment(s) received – characteristics: Type, content, standardisation, frequency, duration (3); specify if different treatments were received (control groups in RCTs…) Describe … any interventions or treatments during this time (D4)			
2.8	Allocation to *treatment*: randomised (NR)/non-randomised (NR) (3)			
	Study design			
2.9	Type of study (cohort, RCT…)			
2.10	Method of recruitment (e.g. consecutive; including whether any stratification or matching was used) (6)			
2.11	Time period (dates/overall duration of recruitment) (6)			
2.12	Start point (6) (e.g. initial presentation to specialist practice…) Were participants enrolled at a similar state of health, or were predictors considered to account for differences? (Ph.3)			
2.13	End point (follow-up): (intended) time of follow-up (6) At what time point was the outcome			

	measured (D3)			
	Outcomes			
2.14	Outcome(s)/definition(s)/measurement(s) (7); specify primary/secondary outcomes where applicable			
2.15	Was/were the same outcome(s)/definition(s)/measurement(s) used as in the original developmental study (where applicable)? Specify any differences			
2.16	Was data provided on (or a reference case made to) the psychometric properties of (the) *outcome measurement(s)* (i.e. when using e.g. available questionnaires, scales, ROM measurements)?			
2.17	If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome (D3)			
2.18	Was a pre-specified outcome definition used? (D3) (specify for each outcome where applicable)			
2.19	Was the outcome defined and determined in a similar way for all participants? (D3) (specify for each outcome where applicable)			
2.20	Was the outcome determined blind to predictor (model) information? (D3) (specify for each outcome where applicable)			
	Prognostic model			
2.21	Prognostic model (or prognostic index/chart) used (including number and list of predictors)			

2.22	Inclusion of predictors which did not use pre-specified or standard predictor definitions (D3.A)				
2.23	Were predictors excluded from the outcome definition? (D3.A)				
2.24	Handling of predictors: any coding/categorisation/dichotomisation? (11) If categorisation of predictors was used, was it pre-specified or justified? (D5)				
2.25	If validation is based on a pre-developed prognostic model (study): Was the very same model with the same coefficients and variables used? Specify any differences with reasons				
2.26	Origin/rationale for used model: (usually) reference to model development study				
2.27	Summary of key aspects of *developmental* study (where applicable/ provided/available)	Population			
		Setting			
		Year			
		Process of predictor selection and inclusion			
		Information on psychometric properties of (the) *predictor measurements*			
		Multivariable analysis methods			
		Other			
		Other			
2.28	Performance statistics of the model used (if based on developmental study): discrimination, validation, any prior				

	validity testing, correction for over-optimism			
2.29	Were predictor assessments blinded to outcome data (D2.A)?			
2.30	Were predictors defined and assessed in a similar way for all participants in the study? (D2.A.)			
2.31	Sample size rationale/target sample size/target power and effect size (where applicable) (9)			
	Statistical and analysis methods			
2.32	Statistical methods used for validation (including e.g. calibration, discrimination…)			
2.33	Pre-specified adjustments/subgroup analyses			
2.34	Handling of missing data (10) (e.g. imputation method) Describe … methods used for missing data (D4)			
3	**Results**			
	Patient flow (12)			
3.1	N of patients meeting inclusion criteria			
3.2	N of dropouts/reasons (12)			
3.3	N of patients selected for inclusion (12)			
3.4	N of dropouts/reasons (12) Describe any participants who were excluded from the analysis (D4)			
3.5	Actual sample size, i.e. *n of patients* included in analysis (for each subgroup where applicable) (12)			
3.6	N of outcome events included in analyses (for each outcome and subgroup where applicable)			
3.7	Actual follow-up time (median/mean)			

	Describe the time interval between predictors and outcome measurement… (D4)				
	Sample characteristics (baseline)				
3.8	Incl. age, sex, types and characteristics of rotator cuff disorder (non-tear, partial or full tears, mixed...), proportions of traumatic disorders, comorbidities, duration of symptoms, baseline functional status) (13)	Age			
		Sex			
		Type/characteristics of disorder (e.g. non-tear, partial or full tears, mixed…)			
		Traumatic origin			
		Duration of symptoms			
		Baseline pain			
		Comorbidities			
		Baseline functional status			
		Dominant arm			
		Employment (e.g. full time, retired…)/work status (sick leave)			
		Other			
		Other			
3.9	Extent of missing values/data (13); Describe missing data on predictors and outcomes… (D4)				
	Analysis and presentation				
3.10	Were all pre-specified outcomes measured and analysed? (D3)				
3.11	Predictors where criteria changed during the study (D2)				
3.12	Details of predictors where data was not provided for all participants (D2)				
3.13	Non constant or subjectively measured predictors where measurement was not blinded to outcome data (D2)				

3.14	Non constant or subjectively measured predictors that were not assessed independently (i.e. without knowledge of each other) (D2)			
3.15	Were all pre-specified outcomes measured and analysed? (D3)			
	Validation statistics			
3.16	Model performance statistics/estimates (including SE, CIs)			
3.17	Discrimination (e.g. C-statistics, AOC) methods and estimates (unadjusted/adjusted) Were the relevant (model) performance measures evaluated (D5)?			
3.18	Calibration (e.g. expected/observed events ratio (CI)) methods and estimates (unadjusted/adjusted) Were the relevant (model) performance measures evaluated (D5)?			
3.19	Were complications in the data (e.g. competing risks, multiple events per individual) accounted for (appropriately)? (D5)			
3.20	Were *non-linear associations* between predictors and the outcome considered and handled appropriately? (D5)			
3.21	Was the model recalibrated or was it stated that recalibration was not needed? (D5)			
3.22	Results of any further investigations, if done (e.g. sensitivity analyses) (18)			
4	**Supplementary notes/information**			

Appendix 3.7

PROBAST version 09/10/2014[17]

Early version • Not intended for use in a review

PROBAST

PROBAST assesses both the risk of bias and applicability of a primary study that aims to develop and/or validate a prediction model.
Bias occurs if systematic flaws or limitations in the design, conduct or analysis of a primary study distort the results. *Risk of bias* refers to the likelihood that a prediction model leads to distorted predictive performance for its intended use and target population. The predictive performance is typically evaluated using calibration, discrimination, and (re)classification.
Applicability refers to the extent to which the model from the primary study matches the review question.

Phase 1. Specify the review question:

Participants (setting, presentation, prior testing, prior treatment):
Outcome to be predicted:
Intended use of model:
Intended moment in time:

Phase 2. Classify the study based on its aim (based on Bouwmeester et al. 2012 PLoS Medicine and TRIPOD 2014):

Type of prediction model study	Tick as appropriate	PROBAST boxes to complete
Prediction model development without external validation. These studies may include internal validation methods such as bootstrapping and cross-validation techniques		Development (Dev) only
Prediction model development combined with external validation in other participants in the same article		Development (Dev) and validation (Val)
External validation of existing (previously developed) model in other participants		Validation (Val) only

[17] The inclusion of this PROBAST version in the thesis appendices was approved by Dr Robert Wolff (personal communication, 17/12/2015).

Phase 3. Risk of bias and applicability judgements

PROBAST is structured as five key domains, each of which is rated in terms of risk of bias (low, high or unclear).

Each domain has a set of signalling questions to help make judgements regarding the risk of bias (rated as yes (Y), probably yes (PY), probably no (PN), no (N) or no information (NI)). All signalling questions are phrased such that "yes" indicates absence of bias. Any signalling question rated as "no" or "probably no" flags the potential for bias, the reviewer then has to use their judgement to determine whether the domain should be rated as "high" or "low" risk of bias. The guidance document contains further instructions on rating risk of bias for each domain.

The first three domains are also rated for concerns regarding applicability (low/ high/ unclear) to the review question (as defined in Phase 1).

DOMAIN 1: Participant selection		
A. Risk of Bias		
(Fill in separately for (i) development (with/without internal validation) and (ii) each external validation dataset) *Describe methods of participant selection:*		
	Dev	Val
Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?		
Were all inclusions and exclusions of participants appropriate?		
Were participants enrolled at a similar state of health, *or* were predictors considered to account for differences?		
Risk of bias introduced by selection of participants **RISK:** *(low/ high/ unclear)*		
Justification of bias rating:		
B. Applicability		
Describe included participants as well as setting and dates of the study:		
Concern that the included participants and setting of the primary study do not match the review question **CONCERN:** *(low/ high/ unclear)*		
Justification of applicability rating:		

DOMAIN 2: Predictors		
A. Risk of Bias		
(Fill in separately for (i) development (with/without internal validation) and (ii) each external validation dataset) *List the predictors assessed:* *List predictors which did not use pre-specified or standard predictor definitions:* *List predictors where criteria for predictors changed during the study:* *List and provide details of predictors where data were not provided for all participants:* *List non constant or subjectively measured predictors where measurement was not blinded to outcome data:* *List non constant or subjectively measured predictors that were not assessed independently (i.e. without knowledge of each other):*		
	Dev	Val
Were predictors defined and assessed in a similar way for all participants in the study?		
Were predictor assessments blinded to outcome data?		
Are all predictors available at the time the model is intended to be used?		
Were all relevant predictors analysed?		
Risk of bias introduced by predictors or their assessment and determination **RISK:** *(low/ high/ unclear)*		
Justification of bias rating:		
B. Applicability		
Concern that the definition, assessment or determination of predictors in the model do not match the intended use of the model specified in the review question **CONCERN:** *(low/ high/ unclear)*		
Justification of applicability rating:		

DOMAIN 3: Outcome		
A. Risk of Bias		
Please complete Domain 3 separately for each outcome (for which a prediction model was developed and/or validated) specified in the review question (Phase 1): *(Fill in separately for (i) development (with/without internal validation) and (ii) each external validation dataset)* *Describe the outcome and how it was defined, assessed and determined:*		
	Dev	Val
Was a pre-specified outcome definition used?[1]		
Were predictors excluded from the outcome definition?		
Was the outcome defined and determined in a similar way for all participants?		
Was the outcome determined blind to predictor information?		
Were all pre-specified outcomes determined and analysed?		
Risk of bias introduced by the outcome or its measurement and determination — **RISK:** *(low/ high/ unclear)*		
Justification of bias rating:		
B. Applicability		
At what time point was the outcome measured: *If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:*		
Concern that the outcomes, their definition, timing or determination do not match the review question — **CONCERN:** *(low/ high/ unclear)*		
Justification of applicability rating:		

[1] Link to COMET and COSIM for pre-specified outcomes

DOMAIN 4: Sample Size and Flow		
Risk of Bias		
(Fill in separately for (i) development (with/without internal validation) and (ii) each external validation dataset) *What was the sample size (number of participants, number of outcome events and number of predictors considered)?* *Describe any participants who were excluded from the analysis:* *Describe the time interval between predictors assessment and outcome determination and any interventions or treatments during this time:* *Describe missing data on predictors and outcomes as well as methods used for missing data:*		
	Dev	Val
Was the number of outcome events per studied predictor reasonable?		
Was the time interval between predictor assessment and outcome determination appropriate?		
Were all enrolled participants included in the analysis?		
Were participants with missing data handled appropriately?		
Risk of bias introduced by sample size or participant flow **RISK:** *(low/ high/ unclear)*		
Justification of bias rating:		

Please complete Domain 5 separately for each model relevant to the review question:

DOMAIN 5: Analysis		
Describe how the model was derived (predictor selection, fitting and optimism, risk groups, model performance): *Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants:*		
Risk of Bias		
	Dev	Val
If categorisation of predictors was used, was it pre-specified or justified?		
Was selection of predictors based on univariable analysis avoided?		
Was optimism in the model performance accounted for, e.g. using bootstrapping or shrinkage techniques?		
Were complications in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?		
Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?		
Were the relevant model performance measures (e.g. calibration, discrimination, (re-)classification, net benefit) of the model (or any simplified score) evaluated?		
Were non-linear associations between predictors and the outcome considered and handled appropriately?		
Was the model recalibrated or was it stated that recalibration was not needed?		
Risk of bias introduced by the analysis **RISK:** *(low/ high/ unclear)*		
Justification of bias rating:		

Phase 4. Reaching an overall judgement about risk of bias

Low risk of bias	If low risk of bias for all domains, the study is judged to be at **low risk of bias** overall.
Moderate risk of bias*	If the study evaluated a prediction model development without external validation, and it was rated as low risk of bias for all domains, consider downgrading to at least **moderate risk of bias**. It is only appropriate for such development study without external validation to be rated as low risk of bias overall if the model was based on a very large data set and included some form of internal validation.
High risk of bias*	The study is judged to be at **high risk of bias** in at least one domain.

* *If an unclear risk of bias is noted for one or more domains, judgement should be used to decide if the study should be classified at a **moderate or high risk of bias** (please refer to the background document for guidance).*

Phase 5. Assess the usability of the model:

Usability		Dev	Val
Are there concerns that the prediction model (including format, presentation and performance) are presented in such a way that it cannot be used in the targeted participants and setting of the review question?	**CONCERN:** *(low/ high/ unclear)*		

Appendix 3.8
PROBAST coding manual

PROBAST Coding Manual (Feb 2015)

Introduction

The PROBAST instrument is completed for each included study, and has five steps. Step 1 requires specification of the review question, and is therefore identical for all included studies. Step 2 requires classification of the study based on its aims, and is self-explanatory.

Step 3, to which this coding manual primarily relates, comprises the Risk of Bias and Applicability judgements for the study. In this context, Risk of Bias is defined as "the likelihood that a prediction model leads to distorted predictive performance for its intended use and target population," and applicability as "the extent to which the model from the primary study matches your systematic review question." Step 3 is divided into five methodological domains, each populated by "signalling" questions.

Step 4 requires an overall judgement about risk of bias and applicability across all assessed domains. Guidance on this rating is provided within the PROBAST instrument. Step 5 requires a judgement about the usability of the model, being defined as "whether the model was presented in enough detail to be usable".

Guidance on answering the signalling questions

Guidance for answering "yes" or "no" is presented on a question-specific basis in Table 1. For most signalling questions (i.e. where not stated differently), there is no question-specific guidance for the other options ("no information", "probably yes" and "probably no"), but the following generic guidance applies.

- Answer "no information" if, due limitations in reporting, the question cannot be answered.
- Answer "probably yes" if it may reasonably be inferred, but is not certain, that the answer is yes.
- Answer "probably no" if it may reasonably be inferred, but is not certain, that the answer is no.

Guidance on rating Summary Risk of Bias at domain level

This guidance is set out in Table 2.

Guidance on rating Summary Applicability at domain level

This guidance is set out in Table 3.

Guidance on rating usability of the model

This guidance is set out in Table 4.

Table 1. Guidance on answering the signalling questions		
DOMAIN 1		
Question	**Guidance**	**Corresponding sections of data extraction forms (DEV & VAL)**
1. Were appropriate data sources used?	Yes if: Cohort, RCT or nested case-control study data was used No if: None of the above sources were used	2.9 (DEV & VAL)
2. Were all inclusions and exclusions of participants appropriate?	Yes if: All inclusions and exclusions matched the study's eligibility criteria No if: There are discrepancies between the inclusions/exclusions and the study's eligibility criteria	2.2, 2.3, 2.8 (DEV & VAL)
3. Were participants enrolled at a similar state of health, or were predictors considered to account for differences?	Note: State of health → consider specifically: duration of complaints, prior shoulder complaints, prior testing or treatment and healthcare setting. Yes if: It is considered sufficiently clear that A) participants were enrolled at a similar state of health or B) that predictors were considered to account for any differences; No if: There are differences regarding the state of health, which are not accounted for (by predictors).	DEV 2.13/ VAL 2.12
DOMAIN 2		
1. Were predictors defined and assessed in a similar way for all participants in the study?	Yes if: This is considered to be satisfied for all predictors; No if: The definition or assessment of one or more predictors differed among the study participants.	DEV 2.28/ VAL 2.30
2. Were predictor assessments made without knowledge of outcome data?	This item is always rated "yes" in the present review, because only prospective studies are included.	DEV 2.27/ VAL 2.29
3. Are all predictors available at the time the model is intended to be used?	This item is always rated "yes" in the present review, because only studies investigating predictors obtained at the baseline assessment are included.	DEV 2.22/ VAL 2.21
4. Were all relevant predictors analysed?	(This question relates to *development* studies only) The question is unanswerable in the present review, because there is insufficient evidence on which to base determination of "all relevant predictors".	
DOMAIN 3		
1. Was a pre-specified outcome definition used?	Yes if: All outcomes are comprehensively specified and defined (complete with any categorization of continuous outcomes, if applicable) *a priori*, in a protocol or the methods section of the study report. No if: Not all outcomes are comprehensively specified and defined, or there appears to be *post hoc* (i.e. data driven) categorisation of any continuous outcome.	DEV & VAL 2.18

2. Were predictors excluded from the outcome definition?	Yes if: A) No predictors were part of the outcome definition or B) one or more predictors were part of the outcome definition, but this is appropriately addressed analytically (e.g. by presenting different models with and without these predictors). No if: A) one or more predictors were part of the outcome definition and this was not appropriately addressed analytically.	DEV 2.26/ VAL 2.23
3. Was the outcome defined and determined in a similar way for all participants?	Self-explanatory	DEV & VAL 2.19
4. Was the outcome determined without knowledge of predictor information?	Yes if: it is explicitly stated that the outcome was determined blind to predictor information No if: it is explicitly stated that the outcome assessment was not blinded Probably yes if: The predictors and outcome included related PROMs. While this is not strictly blind, the patient's judgements are separated by a substantial interval, and so practical blinding may be assumed.	DEV & VAL 2.20
5. Were all pre-specified outcomes determined and analysed?	Yes if: All pre-specified outcomes were determined and analysed by multivariable analysis; No if: Any of the above do not apply.	DEV & VAL 3.15
DOMAIN 4		
1. Were there a reasonable number of outcome events ?	Yes if: There were at least 5-10 outcome events for each predictor[1] included in the multivariable analysis underpinning the reportedly final model or if this is not specified the most complete model including main effects for all prognostic factors No if: The above does not apply	DEV 2.21, 3.6/ VAL 2.21, 3.6
2. Was the time interval between predictor assessment and outcome determination appropriate?	Note: For this review, the 'appropriateness' of the interval takes account of the intervention and its postulated mechanism of effect, and is evaluated case-by-case. Yes if: the duration of treatment, and the relationship between the duration of treatment and duration of follow-up, are clinically reasonable. No if: the duration of treatment, and the relationship between the duration of treatment and duration of follow-up, are not clinically reasonable.	DEV 2.14/ VAL 2.13
3. Were all enrolled participants included in the analysis?	Yes if: It is clear from the study report that A) all participants were included in the analysis, or B) the difference in numbers between enrolled and analysed participants was < 15%[2] No if: the difference is >/= 15%.	DEV 3.1–3.5, 2.35/ VAL 3.1-3.5, 2.34

[1]Vittinghoff, E. & McCulloch, C.E. (2007) Relaxing the rule of ten events per variable in logistic and Cox regression. *American journal of epidemiology*, 165(6), pp.710–8.

[2]Herbert, R.; Jamtvedt, G.; Hagen, K.B.; Mead, J.. (2011) *Practical Evidence-based Physiotherapy,* 2nd ed., Churchill Livingstone.

4. Were participants with missing data handled appropriately?	Yes if: There were missing data, which were addressed transparently and appropriately. (Ideally, but depending on the specific context including the extent and type of missing data, this will involve multiple imputation methods). No if: There were missing data, which were not addressed transparently and appropriately. Not applicable if: The report specifies that there were no missing data.	
DOMAIN 5		
1. Were non-binary predictors handled appropriately	Yes if: A) there was no categorisation of continuous predictors, or B) any categorisation of predictors was specified at the outset (i.e. in a protocol or the methods section of the study report). No if: The study report states that there was *post hoc* (i.e. data-driven) categorisation of any continuous outcome.	DEV 2.31/ VAL 2.24
2. Was selection of predictors based on univariable analysis avoided?	This question relates to *development* studies only; Yes if: The process of selection of predictors for inclusion in the multivariable analysis did not include any automated selection procedure(s) based on the statistical significance of univariable associations between the predictors and the outcome; No if: The process of selecting predictors for inclusion in the multivariable analysis included any automated selection procedure(s) based on the univariable associations between the predictors and the outcome.	DEV 2.30
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	Note: relates to *development* studies only; Yes if: It is reported that model overfitting was assessed and accounted for, irrespective of the applied technique (e.g. bootstrapping, shrinkage); No if: The report specifies that model overfitting was not assessed and accounted for	DEV 3.22
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	Yes if: The occurrence of any complexities is reported, and if these were accounted for; No if: The occurrence of any complexities is reported, and if these were accounted for, but this is considered inappropriate. N/A if: It is explicitly stated that no complexities were observed. Note: This item is considered as of comparably minor relevance within the context of this review, as apart from the possible issue of multiple events per individual (mainly relating to participants versus shoulders), we are unaware of any relevant	DEV 3.23/ VAL 3.19

	complexities, such as competing risks, within the area of interest for this review. This may explain and justify any lack of mention of complexities - (see table 2 for consequences on the overall risk of bias rating for this domain).	
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	Note: This question relates to development studies only; Yes if: The (final) predictors and their assigned weights are presented in addition to the actual results of the multivariable model, and if they correspond; No if: The according details are provided, but if there is a discrepancy between them, i.e. if they appear not to fully correspond. Not applicable if: The (final) predictors and their assigned weights are not presented	DEV 3.17, 3.19
6. For the model or any simplified score, were relevant performance measures evaluated, e.g. calibration, discrimination, (re-)classification, and net benefit?	Yes if: The relevant performance measures (including both discrimination *and* calibration statistics) were evaluated and their statistics were reported. No if: they specify that they haven't done it.	DEV 3.19-3.21/ VAL 3.16-3.18 re-consider
7. Were non-linear associations between predictors and the outcome considered and handled appropriately?	Yes if: A) Non-linear associations between predictors and the outcome were considered through the application of either graphical (residual plots) or statistical methods, and B) the results of these are specified, and C) the results are appropriately dealt with (i.e. if the occurrence of any non-linear associations is appropriately further considered, e.g. by nonlinear transformation No if: they specify that they haven't done it.	DEV 3.24/ VAL 3.20
8. Was the model recalibrated or was it stated that recalibration was not needed?	Note: relates to *validation* studies only: Yes if: The model under investigation was either recalibrated or it was stated that recalibration wasn't needed. No if they specify that they haven't done it.	VAL 3.21

Table 2. Guidance on rating Summary Risk of Bias at domain level	
Rating	**Criterion**
Low	All items within the domain are rated 'yes' or the majority are rated 'yes' with some 'probably yes'.
Unclear	One or more items within the domain are rated 'no information'.
High	One or more items within the domain are rated 'no' or 'probably no'
Note on domain 5	Regarding signalling question 4 ("Were any complexities in the data accounted for appropriately") → The 'unclear' rating for this question will affect the overall risk of bias rating for this domain only if one or more further items were rated as either 'no information' or 'no'.

Table 3. Guidance on rating Applicability at domain level	
Rating	**Criterion**
Low	There are no concerns about applicability.
Unclear	There are moderate concerns about applicability, which may or may not be due to limitations in reporting.
High	There are serious concerns about applicability.

Table 4. Guidance on rating usability of the model	
Rating	**Criterion**
Note	Rate as overall judgement of whether the model(s) is/are “ready for use in clinical practice” considering the phase of research and the outcome, overall judgements of risk of bias, conduct and reporting.
Low	
Unclear	
High	

Appendix 3.9
Detailed results of PROBAST assessment

See Appendix 3.8 for the PROBAST coding manual

Abbreviations and symbols used in the table:

Y = yes; PY = probably yes; N = no; NI = no information; N/A = not applicable; UNANSW = unanswerable (see coding manual); = low risk/concerns; = high risk/concerns (usability: not usable); ? = unclear risk/concerns

PROBAST items & judgements	**Ratings** (see above for explanations)
Hallgren et al. 2014	
Domain 1: Participant selection	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y
2. Were all inclusions and exclusions of participants appropriate?	Y
3. Were participants enrolled at a similar state of health, *or* were predictors considered to account for differences?	Y*
Risk of bias judgement	
Justification of bias rating: *State of health is assumed to be similar: point of failed conservative treatment with referral for surgery (single study centre may be viewed as ensuring consistent approach in selecting participants for surgery) (even though duration of complaints (> 6 months as inclusion criterion) was not considered as a predictor).	
Applicability judgement	
Justification of applicability rating: --	
Domain 2: Predictors	
1. Were predictors defined and assessed in a similar way for all participants?	Y
2. Were predictor assessments blinded to outcome data?	Y
3. Are all predictors available at the time the model is intended to be used?	Y
4. Were all relevant predictors analysed?	UNANSW.
Risk of bias judgement	
Justification of bias rating: --	

PROBAST items & judgements	Ratings (see above for explanations)
Applicability judgement	
Justification of applicability rating: --	
Domain 3: Outcome	
1. Was a pre-specified outcome definition used?	Y
2. Were predictors excluded from the outcome definition?	Y
3. Was the outcome defined and determined in a similar way for all participants?	Y
4. Was the outcome determined blind to predictor information?	Y
5. Were all pre-specified outcomes determined and analysed?	Y
Risk of bias judgement	
Justification of bias rating: --	
Applicability judgement	
Justification of applicability rating: --	
Domain 4: Sample size and participant flow	
1. Was the number of outcome events per studied predictor reasonable?	N†
2. Was the time interval between predictor assessment and outcome determination appropriate?	Y
3. Were all enrolled participants included in the analysis?	Y‡
4. Were participants with missing data handled appropriately?	NI
Risk of bias judgement	
Justification of bias rating: †Number of events/predictors: 41 outcome events (having sugery); appropriateness of the ratio between outcome events and predictors depending on the models, but the model with the largest predictive performance had nine predictor variables: 9 x 5 = 45, so the minimum threshold was not met for this model. ‡Data sheet: 93-97 cases, study report: 95; all fit the defined threshold, though.	
Domain 5: Analysis	
1. If categorisation of predictors was used, was it pre-specified or justified?	Y
2. Was selection of predictors based on univariable analysis avoided?	Y
3. Was optimism in the model performance accounted for, e.g. using bootstrapping or shrinkage techniques?	NI
4. Were complications in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	NI
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	N/A

PROBAST items & judgements	Ratings (see above for explanations)
6. Were the relevant model performance measures (e.g. calibration, discrimination, (re-) classification, net benefit) of the model (or any simplified score) evaluated?	NI
7. Were non-linear associations between predictors and the outcome considered and handled appropriately?	NI
8. Was the model recalibrated or was it stated that recalibration was not needed?	
Risk of bias judgement	?
Justification of bias rating: --	
Overall judgments	
Risk of bias	□
Applicability	□
Usability	□
Hung et al. 2010	
Domain 1: Participant selection	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y
2. Were all inclusions and exclusions of participants appropriate?	PY
3. Were participants enrolled at a similar state of health, *or* were predictors considered to account for differences?	PY§
Risk of bias judgement	□
Justification of bias rating: §With respect to the selection process (see above), and due to lack of (further) information, it seems not completely clear whether the participants were enrolled at a similar state of health, but they probably were, and baseline functional state as well as the duration of symptoms are presented and were (at least initially) considered as predictors.	
Applicability judgement	?
Justification of applicability rating: "SIS diagnosis", but partly unclear (partly lack of details on the actual population, but also on lack of details with eligibility criteria): exclusion of participants with capsulitis-type disorders (and other exclusion criteria such as cervical spine involvement, signs of bony degeneration)? Traumatic cases (unlikely)? Acute inflammation? To be considered: rather young population, males only, i.e. limited representativeness.	
Domain 2: Predictors	
1. Were predictors defined and assessed in a similar way for all participants?	NI
2. Were predictor assessments blinded to outcome data?	Y

□

PROBAST items & judgements	Ratings (see above for explanations)
3. Are all predictors available at the time the model is intended to be used?	Y
4. Were all relevant predictors analysed?	UNANSW.
Risk of bias judgement	?
Justification of bias rating: --	
Applicability judgement	□
Justification of applicability rating: Two predictors, one of which was included in the final model, were not defined. Some concerns about the definition, psychometric properties and (types of) assessments of some of predictors: Limited reproducibility due to partly insufficient specification; doubts specifically about the validity of some of the measurements (e.g. scapular assessments); use of "special equipment" (e.g. FASTREK system) that usually cannot be expected to be available or accessible clearly limits applicability in "everyday clinical practice".	
Domain 3: Outcome	
1. Was a pre-specified outcome definition used?	Y
2. Were predictors excluded from the outcome definition?	Y
3. Was the outcome defined and determined in a similar way for all participants?	Y
4. Was the outcome determined blind to predictor information?	N‖
5. Were all pre-specified outcomes determined and analysed?	Y
Risk of bias judgement	□
Justification of bias rating: ‖Asks for perception of change (from baseline to follow-up). "An independent trained measurer blinded to treatment assessed the outcomes", but it is unclear whether the assessor was blinded to predictor information. Also unclear are the specific scapular kinematic conditions evaluated: these are not comprehensively listed.	
Applicability judgement	?
Justification of applicability rating: Main critique: dichotomisation of outcome with arbitrary cut-off between "improvement" and "non-improvement" tends to potentially limit applicability.	
Domain 4: Sample size and participant flow	
1. Was the number of outcome events per studied predictor reasonable?	Y#
2. Was the time interval between predictor assessment and outcome determination appropriate?	Y
3. Were all enrolled participants included in the analysis?	Y
4. Were participants with missing data handled appropriately?	NI

□?

PROBAST items & judgements	Ratings (see above for explanations)
Risk of bias judgement	?
Justification of bias rating: #See coding manual: in relation to final model.	
Domain 5: Analysis	
1. If categorisation of predictors was used, was it pre-specified or justified?	N**
2. Was selection of predictors based on univariable analysis avoided?	N
3. Was optimism in the model performance accounted for, e.g. using bootstrapping or shrinkage techniques?	NI
4. Were complications in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	NI
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	N/A
6. Were the relevant model performance measures (e.g. calibration, discrimination, (re-)classification, net benefit) of the model (or any simplified score) evaluated?	NI
7. Were non-linear associations between predictors and the outcome considered and handled appropriately?	NI
8. Was the model recalibrated or was it stated that recalibration was not needed?	
Risk of bias judgement	□
Justification of bias rating: **Effectively not. While the method for dichotomising the predictors was pre-specified, this was data-driven.	
Overall judgments	
Risk of bias	□
Applicability	□
Usability	□
Kromer et al. 2014	
Domain 1: Participant selection	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y
2. Were all inclusions and exclusions of participants appropriate?	Y
3. Were participants enrolled at a similar state of health, *or* were predictors considered to account for differences?	PY††
Risk of bias judgement	□

□?

PROBAST items & judgements	Ratings (see above for explanations)
Justification of bias rating: ††Patients referred by general practitioners and those referred by orthopaedic surgeons may be at a different stage of disease. However, predictors were included which accounted for this. (Also, consider "German pathways", which allow for both ways: seeing a GP and seeing a specialist from the outset).	
Applicability judgement	□
Justification of applicability rating: The exclusion criterion > 1/3 restriction of elevation versus the unaffected shoulder lacks sufficient definition: the exclusion of frozen shoulder as diagnosis + the baseline data as provided in the 2013 report p.491, though, suggest that the criteria matched the review question/criteria.	
Domain 2: Predictors	
1. Were predictors defined and assessed in a similar way for all participants?	Y
2. Were predictor assessments blinded to outcome data?	Y
3. Are all predictors available at the time the model is intended to be used?	Y
4. Were all relevant predictors analysed?	UNANSW.
Risk of bias judgement	□
Justification of bias rating: --	
Applicability judgement	□
Justification of applicability rating: --	
Domain 3: Outcome	
1. Was a pre-specified outcome definition used?	Y
2. Were predictors excluded from the outcome definition?	N‡‡
3. Was the outcome defined and determined in a similar way for all participants?	Y
4. Was the outcome determined blind to predictor information?	PY
5. Were all pre-specified outcomes determined and analysed?	Y
Risk of bias judgement	□
Justification of bias rating: ‡‡Incorporation bias/mathematical coupling/RTM, which was not accounted for at the analysis stage	
Applicability judgement	□
Justification of applicability rating: --	

□?

PROBAST items & judgements	Ratings (see above for explanations)
Domain 4: Sample size and participant flow	
1. Was the number of outcome events per studied predictor reasonable?	Y
2. Was the time interval between predictor assessment and outcome determination appropriate?	Y
3. Were all enrolled participants included in the analysis?	Y
4. Were participants with missing data handled appropriately?	NI
Risk of bias judgement	?
Justification of bias rating: --	
Domain 5: Analysis	
1. If categorisation of predictors was used, was it pre-specified or justified?	Y
2. Was selection of predictors based on univariable analysis avoided?	N
3. Was optimism in the model performance accounted for, e.g. using bootstrapping or shrinkage techniques?	NI
4. Were complications in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	NI
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	N/A
6. Were the relevant model performance measures (e.g. calibration, discrimination, (re-)classification, net benefit) of the model (or any simplified score) evaluated?	NI
7. Were non-linear associations between predictors and the outcome considered and handled appropriately?	NI
8. Was the model recalibrated or was it stated that recalibration was not needed?	
Risk of bias judgement	□
Justification of bias rating: --	
Overall judgments	
Risk of bias	□
Applicability	□
Usability	□
Merolla et al. 2011	
Domain 1: Participant selection	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y
2. Were all inclusions and exclusions of participants appropriate?	NI§§

□?

PROBAST items & judgements	Ratings (see above for explanations)
3. Were participants enrolled at a similar state of health, *or* were predictors considered to account for differences?	NI‖‖
Risk of bias judgement:	?
Justification of bias rating: §§Insufficient information on recruitment and inclusion process (+ discrepancy between eligibility criteria and some predictors); ‖‖insufficient information, e.g. no information on duration of complaints (baseline assessment and predictors)	
Applicability judgement	?
Justification of applicability rating: Unclear due to lack of details about the actual sample characteristics.	
Domain 2: Predictors	
1. Were predictors defined and assessed in a similar way for all participants?	NI##
2. Were predictor assessments blinded to outcome data?	Y
3. Are all predictors available at the time the model is intended to be used?	Y
4. Were all relevant predictors analysed?	
Risk of bias judgement	?
Justification of bias rating: ##Lack of operational definitions	
Applicability judgement	□
Justification of applicability rating: Overall very poor definitions of predictors and measurements.	
Domain 3: Outcome	
1. Was a pre-specified outcome definition used?	N***
2. Were predictors excluded from the outcome definition?	PN†††
3. Was the outcome defined and determined in a similar way for all participants?	NI
4. Was the outcome determined blind to predictor information?	NI‡‡‡
5. Were all pre-specified outcomes determined and analysed?	NI‡‡‡
Risk of bias judgement	□
Justification of bias rating: ***No, as the outcome of interest is unclear (see those listet + surgery yes/no + quality of life mentioned somewhere in the text);	

□?

PROBAST items & judgements	Ratings (see above for explanations)
†††CM was used as one outcome (?), which includes (active) ROM, which was a predictor. As it is unclear what outcome(s) were assessed, though, we rated this as PN. ‡‡‡Rated as NI due to lack of information/clarity about assessed outcomes…; see PROBAST coding manual. Not all outcomes are mentioned in the analysis and results section: unclear and confusing report…	
Applicability judgement	?
Justification of applicability rating: Unclear due to poor reporting.	
Domain 4: Sample size and participant flow	
1. Was the number of outcome events per studied predictor reasonable?	PN$$$
2. Was the time interval between predictor assessment and outcome determination appropriate?	NI
3. Were all enrolled participants included in the analysis?	PY
4. Were participants with missing data handled appropriately?	NI
Risk of bias judgement	□
Justification of bias rating: $$$See coding manual. 17 predictors (plus dummy variables!).	
Domain 5: Analysis	
1. If categorisation of predictors was used, was it pre-specified or justified?	Y
2. Was selection of predictors based on univariable analysis avoided?	
3. Was optimism in the model performance accounted for, e.g. using bootstrapping or shrinkage techniques?	
4. Were complications in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	NI
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	
6. Were the relevant model performance measures (e.g. calibration, discrimination, (re-)classification, net benefit) of the model (or any simplified score) evaluated?	NI
7. Were non-linear associations between predictors and the outcome considered and handled appropriately?	NI
8. Was the model recalibrated or was it stated that recalibration was not needed?	NI
Risk of bias judgement	?
Justification of bias rating: Overall: Fair and in line with coding manual to give an “unclear” (?) rating – due to the inappropriate presentation and results it could have been rated as high risk, too…	

□?

PROBAST items & judgements	Ratings (see above for explanations)
Overall judgments	
Risk of bias	□
Applicability	□
Usability	□
Taheriazam et al. 2005	
Domain 1: Participant selection	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y
2. Were all inclusions and exclusions of participants appropriate?	PY||||||
3. Were participants enrolled at a similar state of health, *or* were predictors considered to account for differences?	PY||||||
Risk of bias judgement	□
Justification of bias rating: ||||||See study report (p. 3/6 table 1): 18/89 participants were found to have either moderately impaired (n = 12) or severely impaired (n = 6) *active* ROM. Seems a bit unclear, thus the PY ratings.	
Applicability judgement	?
Justification of applicability rating: see justification of bias rating	
Domain 2: Predictors	
1. Were predictors defined and assessed in a similar way for all participants?	PY###
2. Were predictor assessments blinded to outcome data?	Y
3. Are all predictors available at the time the model is intended to be used?	Y
4. Were all relevant predictors analysed?	UNANSW.
Risk of bias judgement	□
Justification of bias rating: ###Some uncertainty about ROM measurement (lack of details on measurement) as well as few information on some other predictors; seems though, that probably, predictors were defined and assessed in a similar way, i.e. we agreed to give a PY.	
Applicability judgement	?
Justification of applicability rating: Some concerns regarding insufficient definition of some of the predictors. Seems fair, though, to give an intermediate rating = unclear, as for the majority of predictors, these concerns appear not to apply (see predictors).	
Domain 3: Outcome	

□?

PROBAST items & judgements	Ratings (see above for explanations)
1. Was a pre-specified outcome definition used?	Y
2. Were predictors excluded from the outcome definition?	N****
3. Was the outcome defined and determined in a similar way for all participants?	Y
4. Was the outcome determined blind to predictor information?	PY
5. Were all pre-specified outcomes determined and analysed?	Y
Risk of bias judgement:	□
Justification of bias rating: ****Incorporation bias/mathematical coupling/RTM, which was not accounted for at the analysis stage	
Applicability judgement	□
Justification of applicability rating: --	
Domain 4: Sample size and participant flow	
1. Was the number of outcome events per studied predictor reasonable?	Y
2. Was the time interval between predictor assessment and outcome determination appropriate?	Y
3. Were all enrolled participants included in the analysis?	Y
4. Were participants with missing data handled appropriately?	NI
Risk of bias judgement	?
Justification of bias rating: --	
Domain 5: Analysis	
1. If categorisation of predictors was used, was it pre-specified or justified?	Y
2. Was selection of predictors based on univariable analysis avoided?	Y
3. Was optimism in the model performance accounted for, e.g. using bootstrapping or shrinkage techniques?	NI
4. Were complications in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	NI
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	N/A
6. Were the relevant model performance measures (e.g. calibration, discrimination, (re-)classification, net benefit) of the model (or any simplified score) evaluated?	NI
7. Were non-linear associations between predictors and the outcome considered and handled appropriately?	Y
8. Was the model recalibrated or was it stated that recalibration was not	

□?

PROBAST items & judgements	Ratings (see above for explanations)
needed?	
Risk of bias judgement	?
Justification of bias rating: --	
Overall judgments	
Risk of bias	□
Applicability	?
Usability	□

Key: Y = yes; PY = probably yes; N = no; NI = no information; N/A = not applicable; UNANSW = unanswerable (see coding manual); □ *= low risk/concerns;* □ *= high risk/concerns (usability: not usable);* ? *= unclear risk/concerns*

□?

Appendix 4.1
Braun et al. 2013

Due to copyright issues, this thesis version does not include appendices 3.1 and 4.1.

Appendix 4.2
Physiotherapy report form

Schulter-Zentrum • Dr. A. Betthäuser • c/o Ev. Krankenhaus Alsterdorf • Elisabeth-Flügge-Str. 1 •
22337 Hamburg • Tel. 460 735 50 • Email info@schulter-zentrum.com

Insert ID:

Physiotherapy Report

Please complete the questions and return this report to Dr Betthäuser (using the prepared reply envelope). Please make sure that the patient's ID is included in the box in the right top corner of this form. The patient should not be named. Thank you very much for your help!

1. Name of physiotherapy practice: _____

2. Date (day/month/year) of the first and the last physiotherapy session that the patient has received in your practice:

First session (DD/MM/YY): ____/____/____ Last session (DD/MM/YY):____/____ /____

3. Total number of treatment sessions: _____ **4.** Average duration of the treatment sessions: _____ minutes

5. Please tick (put a 'X' in) any of the following elements that you have used while you were treating your patient:

Manual mobilisation techniques (shoulder)	
Soft tissue techniques (shoulder or shoulder girdle)	
Stretching techniques or exercises (shoulder/shoulder girdle)	
Strengthening exercises focused at rotator cuff muscles	
Strengthening exercises focused at shoulder girdle muscles	
Inclusion of high load exercises > 80% max. strength (RM)	
Stabilisation exercises	
Scapula positioning exercises	

Humeral head 'positioning' exercises	
Coordination exercises	
Use of training machines (e.g. pulley)	
Use of small equipment (e.g. elastic bands)	
Provision and supervision of home exercises	
Heat or cold applications	
Anything else? If so, please state:	

6. Did your patient report any problems such as side effects, exacerbations of symptoms or other problems? If so, please note them:

_____

_____

_____

Appendix 5.1

Primary search strategy for identification of factors

Search ID	Search terms*
#1	rotator cuff (MeSH) OR (shoulder AND cuff) OR supraspinatus OR infraspinatus OR teres minor OR subscapularis
#2	(tear OR defect OR impairment OR degeneration OR injury)
#3	partial OR partial-thickness OR incomplete OR bursal side* OR bursal-side* OR joint side* OR joint-side* OR intratend*
#4	conservative treatment OR non-operative treatment OR non-surgical treatment OR physiotherapy* OR physical therap* OR rehabil*
#5	predict* OR determin* OR prognos* and PubMed 'Clinical Queries': 'Clinical Study Category: Prognosis' filter
Example combinations	(#1 and (#2 or #3)) and #4 and #5; #1 and #4 and #5; #1 and #5

*The search terms were used in all possible permutations. The searches were supplemented by "related articles" searches and hand searches of reference lists.

Appendix 5.2
Primary study reports and other articles used to identify prognostic factors

Articles are listed by type - primary (clinical) study, systematic review, expert consensus. The references are also provided in the chapter reference list.

No	Study ID*	Reference
Primary study		
1	Bartolozzi 1994	Bartolozzi A., Andreychik D., Ahmad S. (1994). Determinants of outcome in the treatment of rotator cuff disease. *Clin Orthop Relat Res*, (308), 90-7.
2	Chard 1988	Chard M. D., Sattelle L. M., Hazleman B. L. (1988). The long-term outcome of rotator cuff tendinitis--a review study. *Br J Rheumatol*, 27(5), 385-9.
3	Conroy 1998	Conroy D. E., Hayes K. W. (1998). The effect of joint mobilization as a component of comprehensive treatment for primary shoulder impingement syndrome. *J Orthop Sports Phys Ther*, 28(1), 3-14. doi:10.2519/jospt.1998.28.1.3.
4	Cummins 2009	Cummins C. A., Sasso L. M., Nicholson D. (2009). Impingement syndrome: Temporal outcomes of nonoperative treatment. *J Shoulder Elb Surg*, 18(2), 172-7. doi:10.1016/j.jse.2008.09.005.
5	Ekeberg 2010	Ekeberg O. M., Bautz-Holter E., Juel N. G., Engebretsen K., Kvalheim S., Brox J. I. (2010). Clinical, socio-demographic and radiological predictors of short-term outcome in rotator cuff disease. *BMC Musculoskelet Disord*, 11(1), 239. doi:10.1186/1471-2474-11-239.
6	Engebretsen 2010	Engebretsen K., Grotle M., Bautz-Holter E., Ekeberg O. M., Brox J. I. (2010). Predictors of shoulder pain and disability index (SPADI) and work status after 1 year in patients with subacromial shoulder pain. *BMC Musculoskelet Disord*, 11, 218. doi:10.1186/1471-2474-11-218.
7	Hardy 1986	Hardy D. C., Vogler J. B., White R. H. (1986). The shoulder impingement syndrome: prevalence of radiographic findings and correlation with response to therapy. *AJR Am J Roentgenol,* 147(3), 557-61. doi:10.2214/ajr.147.3.557.
8	Hawkins 1995	Hawkins R. H., Dunlop R. (1995). Nonoperative treatment of rotator cuff tears. *Clin Orthop Relat Res*, (321), 178-88.
9	Hung 2010	Hung C.-J., Jan M.-H., Lin Y.-F., Wang T.-Q., Lin J.-J. (2010). Scapular kinematics and impairment features for classifying patients with subacromial impingement syndrome. *Man Ther*, 15(6), 547-51. doi:10.1016/j.math.2010.06.003.

No	Study ID*	Reference
10	Itoi 1992	Itoi E., Tabata S. (1992). Conservative treatment of rotator cuff tears. *Clin Orthop Relat Res*, (275), 165-73.
11	Kennedy 2006a	Kennedy C. A., Haines T., Beaton D. E. (2006). Eight predictive factors associated with response patterns during physiotherapy for soft tissue shoulder disorders were identified. *J. Clin. Epidemiol.*, *59*(5), 485–96. doi:10.1016/j.jclinepi.2005.09.003.
12	Kennedy 2006b	Kennedy C. A., Manno M., Hogg-Johnson S., Haines T., Hurley L., McKenzie D., Beaton D. E. (2006). Prognosis in soft tissue disorders of the shoulder: predicting both change in disability and level of disability after treatment. *Phys. Ther.*, *86*(7), 1013–1032; discussion 1033–1037.
13	Maman 2009	Maman E., Harris C., White L., Tomlinson G., Shashank M., Boynton E. (2009). Outcome of nonoperative treatment of symptomatic rotator cuff tears monitored by magnetic resonance imaging. *J Bone Joint Surg Am*, 91(8), 1898-906. doi:10.2106/JBJS.G.01335.
14	Morrison 1997	Morrison D. S., Frogameni a D., Woodworth P. (1997). Non-operative treatment of subacromial impingement syndrome. *J Bone Joint Surg Am*, 79(5), 732-7.
15	Safran 2011	Safran O., Schroeder J., Bloom R., Weil Y., Milgrom C. (2011). Natural history of nonoperatively treated symptomatic rotator cuff tears in patients 60 years old or younger. *Am J Sports Med*, 39(4), 710-4. doi:10.1177/0363546510393944.
16	Selvanetti 1998	Selvanetti A., Giombini A., Caruso I. (1998). Nonoperative treatment of partial-thickness rotator cuff tears in overhead athletes. *Med Sci Sports Exerc*, 30(5), S260.
17	Taheriazam 2005	Taheriazam A., Sadatsafavi M., Moayyeri A. (2005). Outcome predictors in nonoperative management of newly diagnosed subacromial impingement syndrome: a longitudinal study. *MedGenMed*, 7(1), 63.
18	Tanaka 2010	Tanaka M., Itoi E., Sato K., Hamada J., Hitachi S., Tojo Y., Honda M., Tabata S. (2010). Factors related to successful outcome of conservative treatment for rotator cuff tears. *Ups J Med Sci*, 115(3), 193-200. doi:10.3109/03009734.2010.493246.
19	Vad 2002	Vad V. B., Warren R. F., Altchek D. W., O'Brien S. J., Rose H. A., Wickiewicz T. L. (2002). Negative prognostic factors in managing massive rotator cuff tears. *Clin J Sport Med*, 12(3), 151-7.
20	Virta 2009	Virta L., Mortensen M., Eriksson R., Möller M. (2009). How many patients with subacromial impingement syndrome recover with physiotherapy? A follow-up study of a supervised exercise programme. *Adv Physiother*, 11(3), 166-73. doi:10.1080/14038190802460481.
21	Wang 2000	Wang J. C., Horner G., Brown E. D., Shapiro M. S. (2000). The relationship between acromial morphology and conservative treatment of patients with impingement syndrome. *Orthopedics*, *23*(6), 557–559.

No	Study ID*	Reference
22	Wu 2003	Wu H. P., Dubinsky T. J., Richardson M. L. (2003). Association of shoulder sonographic findings with subsequent surgical treatment for rotator cuff injury. *J Ultrasound Med*, 22(2), 155-61.
23	Yamanaka 1994	Yamanaka K., Matsumoto T. (1994). The joint side tear of the rotator cuff. A followup study by arthrography. *Clin Orthop Relat Res*, (304), 68–73.
Systematic review		
24	Kuijpers 2004	Kuijpers T., van Der Windt D. A. W. M., van Der Heijden G. J. M. G., Bouter L. M. (2004). Systematic review of prognostic cohort studies on shoulder disorders. *Pain*, 109(3), 420-31. doi:10.1016/j.pain.2004.02.017.
Expert consensus		
25	Vergouw 2011	Vergouw D., Heymans M. W., de Vet H. C., van der Windt D. A., van der Horst H. E. (2011). Prediction of persistent shoulder pain in general practice: Comparing clinical consensus from a Delphi procedure with a statistical scoring system. *BMC Fam Pract*, 12(1), 63. doi:10.1186/1471-2296-12-63.

** First author, year*

Appendix 5.3
Excluded factors (selection stage 2) with reasons

Articles are listed in the same order (and using the same numbers) as in Table 5.1.

1	Factor	Main criterion not met by factor and main reason for exclusion
Factors related to history of symptoms/shoulder pain		
11	Aetiology of symptoms (history of trauma to the shoulder, type of rotator cuff disease and overuse)	Relevance to study □ Considered as largely irrelevant to my study, as inclusion was restricted to PTTs, i.e. a specific type of rotator cuff disease, and as trauma-related tears were an exclusion criterion.
Factors from physical examination		
14	Impingement sign (presence/absence)	Relevance to study □ This was among the clinical signs as part of the diagnostic criteria.
15	Muscle strength (serratus anterior; abduction; rotation)	Applicability/practicability □ Valid measurement would have required equipment that was not available to Dr Betthäuser in his practice (and that is normally not available in the standard practice setting).
16	Range of motion (active: abduction, external rotation; passive)	Relevance to study □ Not considered a relevant issue to the study population. Also, significant restriction of range of movement was an exclusion criterion.
17	Scapular kinematics: internal rotation	Applicability/practicability □ Valid measurement would have required specialist equipment that was not available to Dr Betthäuser in his practice (and that is normally not available in the standard practice setting).
Factors related to comorbidities and (self-reported) health status		
19	Glenohumeral arthritis	Relevance to study □ Glenohumeral arthritis was an exclusion criterion.
21	Multisite pain	Relevance to study □ Multisite pain was an exclusion criterion.
Psychological factors		
23	Fear-avoidance beliefs	Measurement properties□ Non-availability of an appropriate validated German questionnaire.
Structural factors (shoulder)		
26	Acromion type/morphology	Applicability/practicability □ Assessment requires special X-ray images that are not part of standard practice within German statutory healthcare.

□

1	Factor	Main criterion not met by factor and main reason for exclusion
27	Humeral head migration	Measurement properties □ Concerns about validity of ultrasound measurement (insufficient evidence); other imaging techniques are not part of standard practice within German statutory healthcare.
28	Osseous abnormalities (not further specified)	Relevance to study □ Clinically relevant glenohumeral degeneration or disease was an exclusion criterion.
Rotator cuff specific factors		
29	Fatty infiltration	Relevance to study □ Relates mainly to (large) FTTs
30	Muscle atrophy	Relevance to study □ Relates mainly to (large) FTTs
31	Tear size (extent)	Measurement properties □ Concerns about validity of measurement of tear size by ultrasound
32	Type of rotator cuff pathology or tear; tendon integrity	Relevance to study □ Inclusion into the study was restricted to patients with PTTs.
Interventional factors		
33	Corticosteroid injections (response to initial injection; previous)	Applicability/practicability □ Response to initial injection requires all participants to have an injection; this does not comply with standard practice within German statutory healthcare. Validity and reliability of measurement □ Asking about previous injections was considered difficult (participants can't be expected to know what type of injections they have received).
34	Medication (regular medication; over-the-counter medication)	Relevance to study □ Dr Betthäuser's patients (with PTTs) usually do not take any oral pain medication.
Economical factors		
35	Insurance (worker's compensation) claims	Relevance to study □ Dr Betthäuser considered this not to be of any relevance for the study population.
36	Sick leave	Relevance to study □ The patients presenting to Dr Betthäuser are usually not on sick leave.

□

Appendix 5.4
Shoulder PROMs

PROMs are ordered by type of questionnaire. The full references are provided in the chapter reference list.

No	PROM	Abbreviation	Article ID*
General			
Upper extremity-related			
1	Disabilities of the Arm, Hand and Shoulder Questionnaire	DASH (& Quick-DASH)	Bot 2004, Desai 2010, Kirkley 2003, Michener 2001, Roy & Esculier 2011, Roy 2009, Wright 2010
2	Upper Extremity Functional Index (Scale)	UEFI (UEFS)	Bot 2004, Michener 2001, Roy 2011
3	Upper Extremity Functional Limitation Scale	UEFL	Bot 2004, Roy 2011
4	Upper Limb Functional Index	ULFI	Roy 2011
Shoulder-related			
5	American Shoulder and Elbow Surgeons Shoulder Score	ASES	Bot 2004, Habermeyer 2006, Kirkley 2003, Michener 2001, Oh 2009, Roy 2011, Roy 2009, Wright 2010
6	Constant (Shoulder) Score		Habermeyer 2006, Michener 2001, Oh 2009, Wright 2010
7	Oxford Shoulder Score	OSS	Desai 2010, Ekeberg 2008, Habermeyer 2006, Kirkley 2003, Roy 2011
8	Penn Shoulder Score	PSS	Michener 2001, Roy 2011
9	Shoulder Activity Level		Wright 2010
10	Shoulder Disability Questionnaire	SDQ	Bot 2004, Desai 2010, Michener 2001
11	Shoulder Rating Questionnaire	SRQ	Bot 2004, Desai 2010, Habermeyer 2006, Kirkley 2003, Roy 2011
12	Shoulder Severity Index	SSI	Bot 2004, Michener 2001
13	Simple Shoulder Test	SST	Bot 2004, Habermeyer 2006, Michener 2001, Oh 2009, Roy 2011, Roy 2009
14	Shoulder Pain and Disability Index	SPADI	Bot 2004, Desai 2010, Ekeberg 2008, Habermeyer 2006, Michener 2001, Roy 2011, Roy 2009
15	Single Assessment	SANE	Wright 2010

No	PROM	Abbreviation	Article ID*
	Numeric Evaluation		
16	Subjective Shoulder Rating System	SSRS	Michener 2001
17	University of California at Los Angeles Shoulder Score	UCLA	Habermeyer 2006, Kirkley 2003, Oh 2009
Specific			
Disease-specific			
18	Rating Scale for Bankart Repair	ROWE	Habermeyer 2006, Kirkley 2003, Oh 2009
19	Rotator Cuff Quality of Life Index	R-QOL	Bot 2004, Habermeyer 2006, Kirkley 2003, Razmjou 2006, Roy 2011, Wright 2010
20	Western Ontario Osteoarthritis Shoulder Index	WOOS	Bot 2004, Habermeyer 2006, Kirkley 2003, Roy 2011, Wright 2010
21	Western Ontario Rotator Cuff Index	WORC	Ekeberg 2008, Habermeyer 2006, Kirkley 2003, Razmjou 2006, Roy 2011, Wright 2010
Condition-specific			
22	Melbourne Shoulder Instability Scale	MISS	Roy 2011
23	Oxford Shoulder Instability Score		Habermeyer 2006
24	Shoulder Instability Questionnaire	SIQ	Bot 2004, Roy 2011
25	Walch-Duplay Score for instability of the shoulder		Habermeyer 2006
26	Western Ontario Shoulder Instability Index	WOSI	Bot 2004, Habermeyer 2006, Kirkley 2003, Oh 2009, Roy 2011, Wright 2010

** First author, year*

Appendix 5.5
Western Ontario Rotator Cuff Index (WORC)

WESTERN ONTARIO ROTATOR CUFF INDEX (WORC)[1]

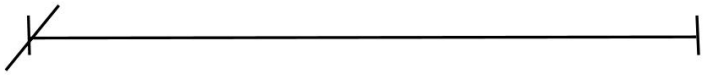
(Kirkely et al. 2003)

INSTRUCTIONS TO PATIENTS

In the following questionnaire you will be asked to answer questions in the following format and you should give your answer by putting a slash "/" on the horizontal line.

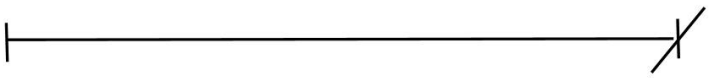
NOTE:

1. If you put a slash "/" at the left end of the line i.e.



then you are indicating that you have no pain.

2. If you put your slash "/" at the right end of the line i.e.



then you are indicating that your pain is extreme.

3. Please note:
a) that the further to the right you put your slash "/", the **more** you experience that symptom.

b) that the further to the left you put your slash "/", the **less** you experience that symptom

c) please do not place your slash outside the end markers.

You are asked to indicate on this questionnaire, the amount of symptom you have experienced in the past week as related to your problematic shoulder. If you are unsure about the shoulder that is involved or you have any other questions, please ask before filling out the questionnaire.

[1] Retrieved/adapted from http://www.shouldercommunity.com/wp-content/uploads/2013/09/Western-Ontario-Rotator-Cuff-Index-WORC.pdf [last accessed 09 June 2016]

If for some reason you do not understand a question, please refer to the explanations that can be found at the end of the questionnaire. You can then place your slash "/" on the horizontal line at the appropriate place. **If an item does not pertain to you or you have not experienced it in the past week, please make your "best guess" as to which response would be the most accurate.**

Section A: Physical Symptoms
INSTRUCTIONS TO PATIENTS

The following questions concern the physical symptoms you have experienced due to your shoulder problem. In all cases, please enter the amount of the symptom you have experienced in the last week. (Please mark your answers with a **slash "/")**

1. How much sharp pain do you experience in your shoulder?

no pain ——— extreme pain

2. How much constant, nagging pain do you experience in your shoulder?

no pain ——— extreme pain

3. How much weakness do you experience in your shoulder?

no weakness ——— extreme weakness

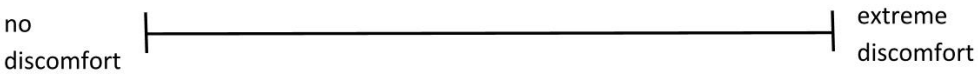
4. How much stiffness or lack of range of motion do you experience in your shoulder?

no stiffness ——— extreme stiffness

5. How much are you bothered by clicking, grinding or crunching in your shoulder?

none ——— extreme

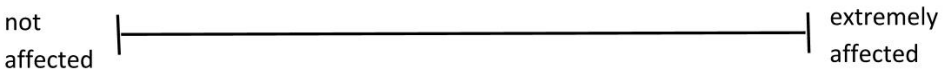
6. How much discomfort do you experience in the muscles of your neck because of your shoulder?



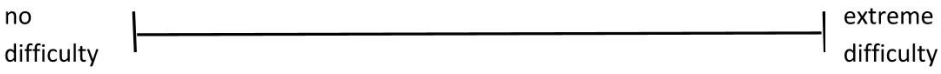
SECTION B: Sports/Recreation
INSTRUCTIONS TO PATIENTS

The following section concerns how your shoulder problem has affected your sports or recreational activities in the past week. For each question, please mark your answers with a slash "/".)

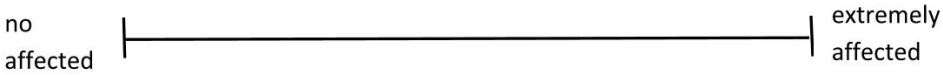
7. How much has your shoulder affected your fitness level?



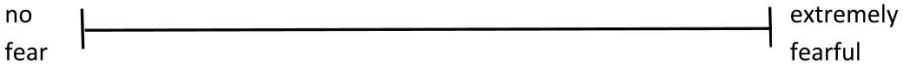
8. How much difficulty do you experience doing push-ups or other strenuous shoulder exercises because of your shoulder?



9. How much has your shoulder affected your ability to throw hard or far?



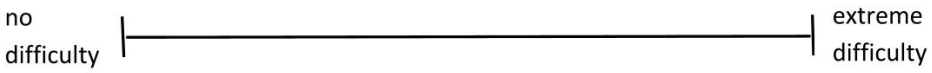
10. How much difficulty do you have with someone or something coming in contact with your affected shoulder?



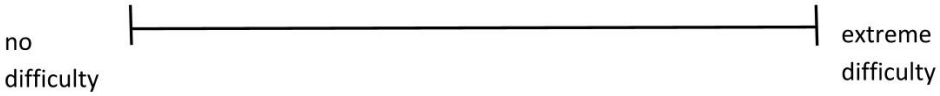
SECTION C: Work
INSTRUCTIONS TO PATIENTS

The following section concerns the amount that your shoulder problem has affected your work around or outside of the home. Please indicate the appropriate amount for the past week with a slash "/".

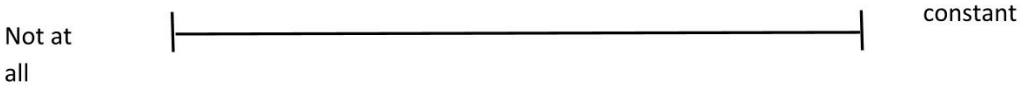
11. How much difficulty do you experience in daily activities about the house or yard?



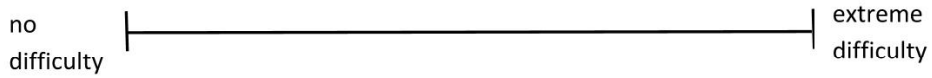
12. How much difficulty do you experience working above your shoulder?



13. How much do you use your uninvolved arm to compensate for your injured one?



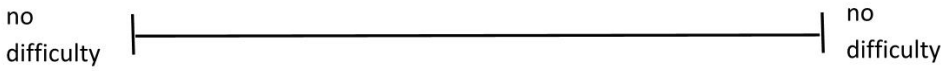
14. How much difficulty do you experience lifting heavy objects at or below shoulder level?



SECTION D: Lifestyle
INSTRUCTIONS TO PATIENTS

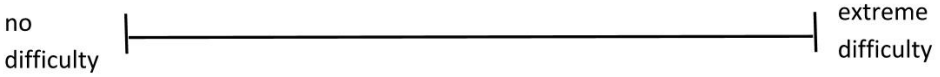
The following section concerns the amount that your shoulder problem has affected or changed your lifestyle. Again, please indicate the appropriate amount for the past week with a slash "/".

15. How much difficulty do you have sleeping because of your shoulder?

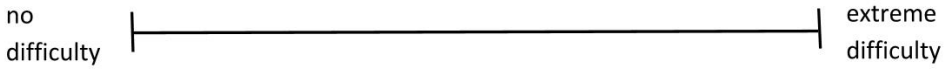


16. How much difficulty have you experienced with styling your hair because of your shoulder?

17. How much difficulty do you have "roughhousing or horsing around" with family or friends?



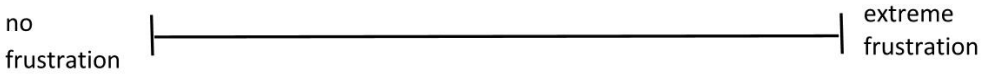
18. How much difficulty do you have dressing or undressing?



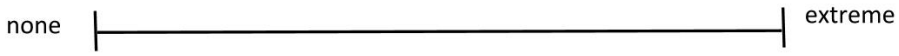
SECTION E: Emotions
INSTRUCTIONS TO PATIENTS

The following questions relate to how you have felt in the past week with regard to your shoulder problem. Please indicate your answer with a slash "/".

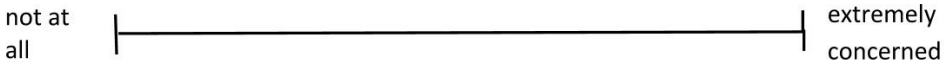
19. How much frustration do you feel because of your shoulder?



20. How much "down in the dumps" or depressed do you feel because of your shoulder?



21. How worried or concerned are you about the effect of your shoulder on your occupation?



THANK YOU FOR COMPLETING THE QUESTIONNAIRE

An Explanation of the Meaning of the Questions in the Western Ontario Rotator Cuff Index (WORK)

Section A: Physical Symptoms

Question 1.
Refers to pain your shoulder that is quick and sudden or that you might refer to as a catching type of pain.

Question 2.
Refers to the dull background ache that always seems to be there as opposed to the sharp pain that is referred to in question 1.

Question 3.
Refers to a lack of strength to carry out a movement.

Question 4.
Refers to the feeling the joint not wanting to move. This is often experienced in the morning upon rising, after exercise or after a period of inactivity. It could also refer to not having full movement of your shoulder in all or any direction(s).

Question 5.
Refers to any of these sounds or feelings that you experience in your shoulder with any type of movement.

Question 6.
Refers to the amount of tension, pain or spasm that you experience in the muscles of your neck that seems to be caused by your shoulder problem.

Section B: Sports/Recreation

Question 7.
Refers to the fitness level you maintained before your shoulder became a problem. Include a decrease in muscle tone or strength level, cardiovascular fitness or strength level.

Question 8.
Refers to any overhead activity requiring you to use some force in its execution. If you do not throw a ball, please consider any other activity such as spiking in volleyball, throwing a stick to your dog, swimming the front crawl, serving in tennis, etc.

Question 9.
Please consider whenever you have been afraid or wary of someone or something hitting or coming into contact with your affected shoulder such as in a sport, a crowded room, an elevator or someone slapping your shoulder in a greeting.

Question 10.
Refers to any exercise requiring you to put force on your shoulder such as push-ups, bench press etc.

Section C: Work

Question 11.
This refers to activities such as raking, shoveling, vacuuming, dusting, weeding, hoeing and washing windows or floors etc.

Question 12.
Refers to any activity requiring you to raise your arms above shoulder level i.e. putting dishes in a cupboard, reaching for an object, painting a ceiling or painting above shoulder level etc.

Question 13.
Refers to if you now use your other arm for any activity or work where you would ordinarily have done it with the arm on the problematic side. If your other shoulder is also symptomatic from Rotator Cuff Disease or some other disease, the consider how you would answer the question if that shoulder was normal.

Question 14.
This does not refer to lifting above your head but to lifting any heavy objects below shoulder level e.g. a bug of groceries, case of pop, suitcase, equipment at work, books, etc.

Section D: Lifestyle:

Question 15.
Refers to having to change your sleeping position, waking up during the night, trouble getting to sleep or waking up feeling unrested.

Question 16.
Refers to anything that you would do to your hair such as combing, brushing or washing that requires you to reach up with your problematic arm.

Question 17.
Refers to any type of rough or vigorous play activity that you would normally engage in with your family or friends.

Question 18.
Refers to reaching behind to do up or undo a zipper or button(s), to do up or undo a bra, pulling on or removing a sweater or top over your head, or tucking in a shirt or top.

Section E: Emotions:

Question 19.
Refers to the frustration you feel because of your inability to do things you used to do or that you want to do but can't.

Question 20.
Down-in-the-dumps or depressed is self-explanatory.

Question 21.
Refers to worrying about your shoulder getting worse instead of better or staying the same and being concerned about what effect that will have on your occupation or work (consider work inside or outside the home).

Appendix 5.6

Pain Catastrophizing Scale (PCS)[18]

PCS

Client No.: _____ Age: _____ Sex: M(__) F(__) Date: _____

Everyone experiences painful situations at some point in their lives. Such experiences may include headaches, tooth pain, joint or muscle pain. People are often exposed to situations that may cause pain such as illness, injury, dental procedures or surgery.

We are interested in the types of thoughts and feelings that you have when you are in pain. Listed below are thirteen statements describing different thoughts and feelings that may be associated with pain. Using the following scale, please indicate the degree to which you have these thoughts and feelings when you are experiencing pain.

0 – not at all **1** – to a slight degree **2** – to a moderate degree **3** – to a great degree **4** – all the time

When I'm in pain …

1. ☐ I worry all the time about whether the pain will end.
2. ☐ I feel I can't go on.
3. ☐ It's terrible and I think it's never going to get any better.
4. ☐ It's awful and I feel that it overwhelms me.
5. ☐ I feel I can't stand it anymore.
6. ☐ I become afraid that the pain will get worse.
7. ☐ I keep thinking of other painful events.
8. ☐ I anxiously want the pain to go away.
9. ☐ I can't seem to keep it out of my mind.
10. ☐ I keep thinking about how much it hurts.
11. ☐ I keep thinking about how badly I want the pain to stop.
12. ☐ There's nothing I can do to reduce the intensity of the pain.
13. ☐ I wonder whether something serious may happen.

…Total

[18] Retrieved from: http://sullivan-painresearch.mcgill.ca/pcs.php [Last accessed 11 June 2016]

Appendix 6.1

Teesside University ethics approval letter

Inspiring success

Teesside University

PRIVATE AND CONFIDENTIAL

Direct Line: 01642 384124

23rd May 2012

Nigel Hanchard
School of Health & Social Care
Teesside University

Dear Nigel

Study No 080/12 - Predicting the outcome of conservative treatment with physiotherapy for shoulder pain in the presence of atraumatic partial-thickness tears of the rotator cuff. Researcher: Cordula Braun. Supervisor: Nigel Hanchard.

Decision: Approved with Advisory Comments

Thank you for your application to the School of Health & Social Care Research Governance and Ethics Committee. The Committee reviewed and approved your application on 15th May 2012 and your study may proceed as it was described in your application pack, with the following advisory comments:

Advisory
Section 13a and throughout: It is not secure to store data on an individual personal computer - please ensure that all study data is stored on the student's U-Drive to take advantage of the University's comprehensive security and backup systems. Please amend all references to this aspect of the study (particularly on the public documents such as PIS and Consent Form) to state '*…data will be stored on a password protected Teesside University server.'*
Participant Information Sheet and all other documents: Please change the phrasing throughout the application pack - and in particular the public documents - from “we” to “I”, "us" to "me" etc as this is as this is currently presented as an individual’s, and not a research team’s project. If this is actually a research team’s project please amend all sections and documents to clarify that. Please give the names, job titles and employers for all members of the research team and detail each person's role and tasks in the project. Please clearly detail the line management relationships and responsibilities from Chief Investigator down through the team.
Consent Form: Please include a *Witness signature* line.

School of Health & Social Care

Inspiring success

Teesside University

Please note:
Where applicable, your study may only proceed when you have also received written approval from any other ethical committee (e.g. NRES) and operational / management structures relevant (e.g. Local NHS R&D). A copy of this approval letter **must** be attached to applications to any other ethical committee. If applicable please forward to me a copy of the approval letter from NRES before proceeding with the study.

In all cases, should you wish to make any substantial amendment to the protocol detailed, or supporting documentation included, in your approved application pack (other than those required as urgent safety measures) you must obtain written approval for those, from myself and all other relevant bodies, prior to implementing any amendment. Details of any changes made as urgent safety measures must be provided in writing to myself and all other relevant bodies as soon as possible after the relevant event; the study should not continue until written approval for those changes has been obtained from myself and all other relevant bodies.

On behalf of the School of Health & Social Care Research Governance and Ethics Committee please accept my best wishes for success in completing your study.

Yours sincerely

Dr. Alasdair MacSween

Chair
Research Governance and Ethics Committee
School of Health & Social Care

School of Health & Social Care

Appendix 6.2

Hamburg Medical Council ethics approval letter

Cc.: Frau Braun per Mail: cordula_braun@gmx.de

ETHIK-KOMMISSION DER
ÄRZTEKAMMER
HAMBURG
Körperschaft des öffentlichen Rechts

Ärztekammer Hamburg · Postfach 76 01 09 · 22051 Hamburg

Herrn
Dr. med. Andreas Betthäuser
Schulter-Zentrum Hamburg
Erste Brunnenstraße 1
20459 Hamburg

09.11.2012

Bearb.-Nr.: **PV4154 (Bitte stets angeben!)**
Studientitel: Prognostische Faktoren für das Ergebnis konservativer Therapie mit Physiotherapie für Patienten mit Schulterschmerzen und nicht-traumatischen partiellen Rotatorenmanschettenrupturen

Sehr geehrter Herr Dr. Betthäuser,

den Eingang Ihres Schreibens vom 04.10.2012 mit Ihrer Stellungnahme und den darin enthaltenen revidierten Studienunterlagen bestätigen wir hiermit.

Die Auflagen der Ethik-Kommission sind nunmehr erfüllt, ein zustimmendes Votum kann somit erteilt werden (Anlage).

Mit freundlichen Grüßen

Antje Diène

i.A. Antje Diène
(Sachbearbeitung)

Humboldtstraße 67a · 22083 Hamburg
Telefon 040/20 22 99-240 · Fax 040/20 22 99-410
ethik@aekhh.de · www.aerztekammer-hamburg.de

Bankverbindung:
Deutsche Apoth. u. Ärztebank, BLZ 300 606 01, Konto-Nr. 000 1346 113

Cc.: Frau Braun per Mail: cordula_braun@gmx.de

ETHIK-KOMMISSION DER
ÄRZTEKAMMER
HAMBURG
Körperschaft des öffentlichen Rechts

Ärztekammer Hamburg · Postfach 76 01 09 · 22051 Hamburg

Herrn
Dr. med. Andreas Betthäuser
Schulter-Zentrum Hamburg
Erste Brunnenstraße 1
20459 Hamburg

09.11.2012

Bearb.-Nr.: PV4154 (Bitte stets angeben!)
Studientitel: Prognostische Faktoren für das Ergebnis konservativer Therapie mit Physiotherapie für Patienten mit Schulterschmerzen und nicht-traumatischen partiellen Rotatorenmanschettenrupturen

Sehr geehrter Herr Kollege Betthäuser,

über Ihr oben bezeichnetes, zur Primärberatung vorgelegtes Projekt hat die Ethik-Kommission ausführlich beraten.

Das Vorhaben entspricht den berufsrechtlichen bzw. gesetzlichen Anforderungen. Die Ethik-Kommission stimmt dem Vorhaben zu.

Die Kommission weist darauf hin, dass die Verantwortung des Versuchsleiters für das Forschungsvorhaben und seine Durchführung durch das obige Votum der Kommission nicht berührt wird.

Für den Fall der Durchführung der Studie in Zentren anderer Kammerbereiche geht die Kommission von der Einbindung der lokal zuständigen Ethik-Kommmission aus.

Sie werden gebeten, die Ethik-Kommission über alle schwerwiegenden oder unerwarteten Ereignisse, die während der Studie auftreten und die die Sicherheit der Studienteilnehmer gefährden, in Verbindung mit Ihrer Stellungnahme zu unterrichten.

Die Kommission geht davon aus, dass die personenbezogenen Daten der Probanden / Patienten den datenschutzrechtlichen Vorschriften entsprechend behandelt werden.

...2

Bankverbindung:
Deutsche Apoth. u. Ärztebank, BLZ 300 606 01, Konto-Nr. 000 1346 113

Humboldtstraße 67a · 22083 Hamburg
Telefon 040/20 22 99-240 · Fax 040/20 22 99-410
ethik@aekhh.de · www.aerztekammer-hamburg.de

Bearb.-Nr.: **PV4154 (Bitte stets angeben!)**
Studientitel: Prognostische Faktoren für das Ergebnis konservativer Therapie mit Physiotherapie für Patienten mit Schulterschmerzen und nicht-traumatischen partiellen Rotatorenmanschettenrupturen

Die Ethik-Kommission erwartet, dass ihr nach Abschluss des Projektes unaufgefordert ein Abschluss-Bericht übersandt wird (unter Angabe der Bearb.-Nr.), aus dem der Erfolg/Misserfolg der Studie sowie Angaben darüber, ob die Studie abgebrochen oder geändert bzw. ob Regressansprüche geltend gemacht wurden, ersichtlich sind.

Mit freundlichen Grüßen
im Auftrage der Kommission

Prof. Dr. med. F. U. Beil
- Stellv. Vorsitzender -

P.S. Die Ethik-Kommission arbeitet auf der Grundlage deutschen Rechts und Berufsrechts sowie in Anlehnung an die ICH-GCP.

PhD Cordula Braun, D4147576

Translation of letter and the approval document from the Ethics Commission of the Medical Council Hamburg, 09 Nov 2012:

---

1. Text of letter (first page):

Processing Nr.: **PV4154 (Please quote at all times!)**
Study title: (English title): "Predicting the outcome of conservative treatment with physiotherapy for shoulder pain in the presence of atraumatic partial-thickness tears of the rotator cuff"

Dear Dr. Betthäuser,

We herewith confirm the receipt of your letter from 10/04/2012 with your comments and the included revised study documents.

The requirements of the ethics commission are now fulfilled, an approving vote can thus be granted (appendix).

With kind regards,
...

---

2. Text of approval document (second and third page):

Processing Nr.: **PV4154 (Please quote at all times!)**
Study title: (English title): "Predicting the outcome of conservative treatment with physiotherapy for shoulder pain in the presence of atraumatic partial-thickness tears of the rotator cuff"

Dear colleague Dr. Betthäuser,

The ethics commission has thoroughly counseled the project that you submitted for primary consultation.

The project complies with the requirements of the rules of professional conduct and with the legal requirements. The ethics commission approves the project.

The commission points out that the responsibility of the principal investigator for the research project and its conduct is not affected by the above vote of the commission.

PhD Cordula Braun, D4147576

In the case of conduct of the study in centers belonging to other council areas, the commission assumes the involvement of the local responsible ethics commission.

We ask you to inform the ethics commission, in connection with your statement, on any serious or unexpected events that occur in the course of the study and that compromise the safety of the study participants.
The commission assumes that the personal data of the participants/patients are treated according to the data protection regulations.

The ethics commission expects to be sent a final report (unasked) after the end of the projct (under quotation of the processing number), from which the following will be evident: the success/failure of the study as well as information on whether the study was cancelled or changed, or whether claims to recourse were alleged.

With kind regards,
On behalf of the Commission.

Prof. Dr.med. F.U. Beil
- Deputy Chair

P.S.: The ethics commission works on the foundation of German law and preofessional law as well as in following the ICH-GCP.

Appendix 6.3
Prognostic study protocol

Prognostic model study protocol

(November 2012)[1]

Note: This document originally had stand-alone appendices. These have now been moved to the respective chapter appendices, and references to them have been relabelled.

1. Study Title

Predicting the outcome of conservative treatment with physiotherapy for shoulder pain in the presence of atraumatic partial-thickness tears of the rotator cuff.

2. Background

The rotator cuff, a deep cuff of four tendons around the shoulder, is often involved in a degenerative continuum culminating in tears (Cook & Purdam 2009; Lewis 2010). The prevalence of these tears has been reported as > 40% (Reilly et al., 2006) and is strongly associated with ageing (Beaudreuil et al., 2007). Rotator cuff tears can significantly affect shoulder function and health-related quality of life (Piitulainen et al., 2012). This underlines their relevance for clinical practice and research.

Rotator cuff tears may be partial-thickness or full-thickness. This study will focus on PTT, as the greatest uncertainty surrounds their treatment. Generally, conservative treatment including physiotherapy is tried before considering surgery (Beaudreuil et al., 2010; Finnan & Crosby, 2010; Smith & Smith, 2010), but this practice lacks a sound evidence basis. The available research reflects the current uncertainty as to precise indications for different treatment approaches: some patients with PTTs respond to conservative treatment including physiotherapy (Ainsworth & Lewis, 2007; Huisstede et al., 2011; Seida et al., 2010), while others respond to operative treatment.

Early identification of likely responders and, by corollary, non-responders to conservative treatment could save effort and suffering and could promote the optimal distribution of available resources. This study aims to address the uncertain indications for the conservative treatment of PTT.

[1] As approved by Teesside University's School of Health & Social Care RG&E Committee (May 2012) and the ethics committee of the Medical Council Hamburg (November 2012).

3. Study aims and research question

Primary aim

To develop a predictive model for the outcome of conservative treatment with physiotherapy for patients with shoulder pain and ultrasonographically diagnosed, atraumatic, partial-thickness tears of the rotator cuff.

The focussed research question is: what combination of factors can best predict the outcome of a course of physiotherapy (with or without adjunctive medical treatment) in patients with shoulder pain and ultrasonographically diagnosed, atraumatic PTT?

Secondary aims

- To determine a 'Minimal clinically important difference' (MID) estimate for the patient-reported outcome measure which is in use with this patient population.
- Using routinely collected data, to document the prevalence of tear progression, and to document the occurrence of physiotherapy-related adverse events.

4. Study design

This study is a quantitative, prospective, non-experimental (observational) study. The design is a single-group cohort study.

5. Study setting and participants

Participants will be patients presenting to an orthopaedic specialist (Dr Andreas Betthäuser) in Hamburg, Germany, with shoulder pain, and who, based on clinical and ultrasonographic examination, are diagnosed with atraumatic partial-thickness tears of the rotator cuff. The physiotherapy treatment (see further) will take place in collaborating physiotherapy practices in the wider area of Hamburg, Germany.

The complete *eligibility criteria* are as follows:

Inclusion:

- Patients with (local) shoulder pain in the presence of an atraumatic (ultrasonographically detected) partial thickness rotator cuff tear (PTT)

- Clinical signs of shoulder impingement (e.g. painful arc, positive impingement signs (e.g. Hawkins-Kennedy))
- Adults (≥ 18 years)
- No restrictions on gender
- Agreement on conservative (i.e. non-surgical) treatment
- Ability to speak and comprehend the German language
- Agreement to participate (signed informed consent)
- Anticipated availability for follow-up (living in area of Hamburg)
- Agreement to physiotherapy in one of the collaborating practices.

Exclusion:

- Presence of a full thickness rotator cuff tear (FTT) at the affected shoulder
- Previous substantial shoulder trauma (e.g. shoulder dislocation)
- Previous surgery for the affected shoulder
- Previous surgery in the shoulder area that may be causal of or contributory to the current problem (e.g. surgery for breast cancer)
- Clinical or (if available) radiological evidence of structural joint pathology
- Significantly restricted passive range of movement (ROM) at the affected shoulder ('capsulitis-type disorders'); current shoulder infection
- Clinical signs of symptomatic acromioclavicular arthritis (e.g. local tenderness, positive provocation tests, e.g. 'Cross-Body Adduction Stress' test)
- Calcific tendinitis
- Ultrasonographic evidence of Long Head of Biceps (LHB) tendon subluxation/ dislocation
- Referred pain from the cervical spine region
- 'Multisite musculoskeletal pain'
- Systemic diseases or comorbidities as potential sources of (the current) shoulder pain (e.g. breast cancer, rheumatoid disease), or as impairing treatment (e.g. cancer, cardiac insufficiencies)
- Neurological disorders or deficits as potential sources of (the current) shoulder pain or impairing assessment and treatment (e.g. hemiplegic shoulder)
- Worker's compensation claims
- Unwillingness or inability to give informed consent (e.g. cognitive or intellectual impairments)

6. Treatment

Participants will be treated conservatively with physiotherapy over a period of three months. Physiotherapy treatment will be based on a broad treatment protocol, which is based on the findings from a review of the literature on physiotherapy interventions (manual therapy and exercises) for impingement-related shoulder pain. If needed, participants may also receive adjunctive medical treatment (such as pain medication or subacromial corticosteroid injections). For further details on the treatments are given with the study procedures.

7. Outcome measures

Outcome variables will include the Western Ontario Rotator Cuff Index (WORC; change from baseline to follow-up (Appendix 5.5), self-perceived change (GPC scale, Appendix 6.9) and progression of the rotator cuff tear (measured by ultrasonography). Further, physiotherapy-related adverse events will be documented. The GPC scale will also be used as the anchor for the determination of an estimate of the MID of the WORC (see study aims).

8. Prognostic factors

The following ten candidate prognostic factors will be investigated:

- Age,
- Sex,
- Pain
- Disability (health-related quality of life, WORC)
- Duration of symptoms
- History of shoulder pain
- Pain catastrophizing
- Physical demands
- Diabetes
- Smoking

9. Sample size estimation and expected duration of data collection

The aim is to derive a prognostic model to predict “success” of treatment (improvement beyond the MID). Due to the multivariable nature of prognostic development studies, it is

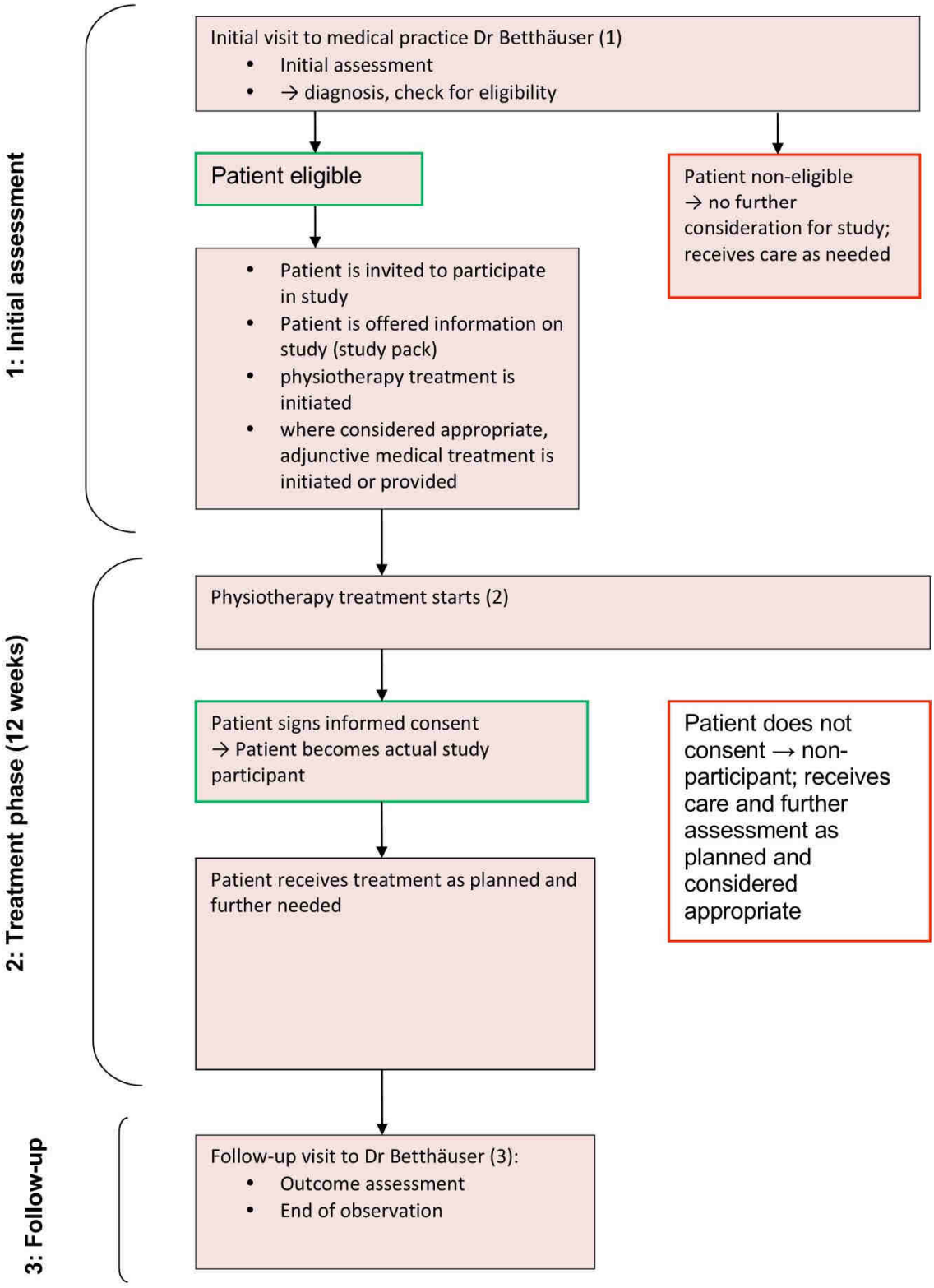
difficult to estimate the required sample size (Moons et al., 2009). Indeed, there are no formal methods (based on either power calculations or adequate precision of estimation of effects) to determine the effective sample size. A commonly used rule of thumb is that there should be at least 10 events per candidate predictor variable (Peduzzi et al., 1996), with the number of events relating to the smallest group (in this case, the proportion of treatment successes or failures) (Bouwmeester et al., 2012). However, more recently it has been proposed that this guideline might be too conservative, and fewer than 10 events per predictor might provide adequate accuracy and precision of estimation of effects. In a comprehensive and rigorous simulation study Vittinghoff and McCulloch (Vittinghoff & McCulloch, 2007) reported that confidence interval coverage, Type 1 error rate, and relative bias were comparable in scenarios with 5-9 events per predictor variable versus those with 10-16 events per variable. Following this work, we base our minimum sample size on a requirement for 5 events per candidate variable. We assume that the proportion of participants benefiting from treatment (treatment successful) is 67%; that is, two-thirds of participants benefit. The smallest group (treatment failures) is therefore a proportion of 0.33. With 10 candidate predictors in our study, the required minimum sample size is therefore 5*10/0.33 = 152 participants. Allowing for a loss to follow-up of 20% the required sample size is 190 participants (152/0.8); our recruitment target is 200 patients.

The expected duration of the data collection period is up to 3 years.

10. Study procedures

Figure 1 provides an overview of the complete planned study process.

Figure 1: Overview of study procedures



10.1. Recruitment phase

10.1.1 As a routine, patients presenting to Dr Betthäuser with shoulder pain will receive an assessment by him including:

a) A subjective assessment, which incorporates patient-reported outcome measures
b) A clinical examination
c) An ultrasound assessment of both shoulders

On the basis of this assessment, and upon fulfillment of the defined eligibility criteria (see section 4), Dr Betthäuser will identify potential study participants.

10.1.2. Dr Betthäuser will have ‘study packs’ to allocate. Each study pack will be in a plastic case labeled with a unique ID and the statement, “Please take this pack with you to your first physiotherapy appointment, even if you decide not to take part in the study”, and will contain:

a) A Patient Information Sheet (PIS, Appendix 6.10), labeled with the unique ID
b) An Informed Consent form (Appendix 6.11), without the unique ID.
c) A list with the names and contact details of the collaborating physiotherapy practices (not appended for reasons of confidentiality
d) A physiotherapy report form (Appendix 4.2), labeled with the unique ID

10.1.3. If Dr Betthäuser decides a patient is eligible, he will

a) Tell the patient about the study, and will invite him/her to consider participating.

For those who choose to consider participating, Dr Betthäuser will

a) Insert the physiotherapy prescription into a pack, and give it to the patient.

b) Transcribe the unique ID onto a coding list, which will be stored separately from the other documents in a locked cabinet in Dr Betthäuser’s practice.

c) Arrange a follow-up assessment at three months after the initial presentation (as is his standard practice). Patients who do not want to fix this appointment immediately will be advised to arrange it as soon as possible.

d) Initiate/provide further medical treatment (e.g. oral medication, local injections to the shoulder), if appropriate.

10.1.4. All eligible patients who are interested in the study will leave Dr Betthäuser's practice with a 'study pack' including their physiotherapy prescription. They will be invited to read the information and informed that they can contact me (or Dr Betthäuser) in case of any further questions about the study. They will further be asked to keep the pack together with their prescription, and to take it with them when they see their physiotherapist irrespective of whether they decide to participate.

Patients wishing to participate will choose one of the collaborating physiotherapy practices for their treatment and, to allow sufficient time for them to decide whether to participate in the study, they will be told that if they decide to participate they should not sign their consent form until they see their physiotherapist.

10.2. Treatment phase

10.2.1. Physiotherapists at the collaborating practices will ask all patients coming from Dr Betthäuser's practice whether or not they have been invited to the study. Those who have been invited will be asked whether or not they have decided to participate.

a) If they have not been invited, or if they do not wish to participate, no further mention will be made of the research.
b) Patients who have been invited and wish to participate will be asked to hand the signed consent form to their physiotherapist. The physiotherapist will put the practice's name on the form. This is to enable me to contact the practices in case of any enquiries or in case of a participant's decision to withdraw.
c) The physiotherapist will send the signed consent form to Dr Betthäuser's practice; using the prepared reply envelope)

10.2.2. Potential participants will become actual participants on receipt of their signed consent forms by.

10.2.3. Patients might have decided to participate but forget to bring their study pack or any of the documents with their unique ID with them when they come to see their physiotherapist for the first time. Thus, there may be cases where the physiotherapist does not know the patient's unique ID. This will be dealt with as follows:

a) The physiotherapist will add a note to the patient's documentation that the unique ID is missing. The patient will be asked to sign a spare consent form. The

physiotherapist will add a note that the unique ID is missing and send the form to Dr Betthäuser (using a reply envelope).

b) Following receipt of the consent form, I will retrieve the patient's unique ID from the coding list (see section 6.1.3). I will then contact the relevant practice, either by phone or by a scheduled visit, to provide the treating physiotherapist with the ID. The physiotherapist will then attach it to the patient's documentation as described above.

10.2.4. The treating physiotherapist will add a note to the patient documentation when a patient has given consent. The physiotherapist will also attach the report form containing the unique ID to the patient documentation, ready for completion at the end of the course of treatment. Where documentation is done electronically, the physiotherapist will document participation, transcribe the unique ID into the database and keep the report form separately. Where a patient has forgotten to bring the study pack and the report form with him/her, the physiotherapist will use a spare report form.

10.2.5. No procedures will be carried out on participants which are not part of their routine care, and only routinely collected data will be utilised.

10.2.6. Participants will be advised on the number and type of follow-up prescriptions for physiotherapy on an individual basis by their treating physiotherapist, and in agreement with Dr Betthäuser. The type, content or amount of treatment will be in no way regulated for the purpose of this study. Thus, the amount and content of treatment may vary among the participants. Details will be documented. Any assessments and treatments will be delivered in accordance with national healthcare or insurance regulations. The follow-up after three months reflects Dr Betthäuser's standard practice, and is not intended to imply completion of a specific number of physiotherapy sessions.

10.2.7. Patients may choose to see Dr Betthäuser during the treatment phase if they feel in need of further advice or adjunctive medical treatment. Dr Betthäuser will document such consultations as well as any treatment he may give (as is usual practice) in his database, so that these will be accessible there.

10.2.8. When the physiotherapy treatment phase ends (after approximately 10-12 weeks), the physiotherapist will

a) Remind the patients of the follow-up appointment with Dr Betthäuser.

b) Complete the report form and send it back to Dr. Betthäuser.

The act of doing this will remove the physical link between the patient's unique ID and his or her physiotherapy documentation. In those physiotherapy clinics where documentation is done electronically, the physiotherapist will delete the ID from the patient database.

10.2.9. There may be situations in which the physiotherapist fails to complete and send back the report. I will monitor the receipt of the reports. If none is received by three months after receipt of the consent form, I will contact the practice.

10.3. Follow-up

10.3.1. The timing and content of follow-up consultations with Dr Betthäuser will not deviate from his usual practice. Thus they will take place approximately 3-4 months after the initial assessment and include:

a) A subjective assessment, incorporating patient-reported outcome measures,
b) A clinical examination
c) A follow-up ultrasound scan of the affected shoulder

10.3.2. There may be cases where patients do not turn up for follow-up, or where they, for any reason, fail to have their assessment within 3-4 months after signing the consent form). For the purposes of the study, this will be dealt with as follows. Patients who have not had their follow-up assessment by 12 weeks after their consent will be sent a letter, inviting them to make another appointment with Dr Betthäuser (provided with the Teesside ethics application). This letter will include the standard follow-up assessment questionnaires, each labeled with the unique ID and a reply envelope (pre-stamped/pre-addressed to Dr Betthäuser). Patients who do not want to or are unable to make another appointment will be invited to complete and return the questionnaires by post.

10.3.3. Dr Betthäuser stores his clinical data in a combination of paper (completed questionnaires) and electronic media. I will abstract all of the relevant data, labeled by ID and not name, onto a study database, which will be kept on a password protected Teesside University server The follow-up assessment marks the end of the observation and of the involvement of the patient as a participant in this study.

All participants will be asked to give consent (on the PIS) to be possibly contacted for participation in a further study (subject to necessary ethical approvals; see PIS (Appendix

6.10) and consent form (Appendix 6.11). The unique ID of all participants will be deleted from the patient's record on Dr Betthäuser's clinical database before data analysis begins (as requested by the ethics commission of the Hamburg Medical Council).

11. Statistical analysis

The analyses will be advised and supervised by Prof Alan Batterham. The primary analysis for the development of the prognostic model will be a multivariable regression analysis. The secondary analysis will be a quantitative anchor-based approach to the determination of a 'Minimally important difference' (MID) estimate for the primary outcome measure (the WORC).

12. Research integrity

12.1. Research standards

This research study was planned in accordance with the principles of the

b) Declaration of Helsinki (World Medical Association, 2008)
c) German Good Clinical Practice Standards (GCP-Verordnung, 2004) and
d) Teesside University's Guidelines for Good Conduct in Research (University of Teesside, 2007).

All data will be treated strictly confidentially and in accordance with the legal regulations on data protection: a) (Data Protection Act, 1998), b) German Federal Data Protection Act (Bundesdatenschutzgesetz, 2009), c) Hamburg Data Protection Act (Hamburgisches Datenschutzgesetz, 1990).

12.2. Confidentiality and Data protection issues

The risk associated with data collection or storage is considered minimal, as

a) The participants' data will be consolidated and stored in Dr Betthäuser's practice. It will be stored on paper format as well as on storage media in a locked cabinet.
b) The key to the personal data, i.e. the coding list for the pseudo-anonymisation, will be kept only at one place, in Dr Betthäuser's practice, separate from the pseudo-anonymised study data, and in a locked cabinet. Access to this list will, besides Dr Betthäuser, be restricted to me.

c) All patient data that will leave either Dr Betthäuser's or the treating physiotherapist's practices will be pseudo-anonymised, i.e. all names will be replaced by a unique code number.
d) All paper-based documents will be stored securely in Dr Betthäuser's practice, as is his responsibility as the treating doctor. Dr Betthäuser holds an electronic patient database that is only accessible through a secure password. I will be in possession of the password operant for the period of data collection.
e) Questionnaires, which Dr Betthäuser maintains in a paper database, will be abstracted by me into electronic format and stored on my research database on a password protected Teesside University server.
f) Follow-up questionnaires returned by post will be labelled by unique ID only. These will be abstracted onto my research database. IDs will be removed before the documents are stored with Dr Betthäuser's records.
g) All data on the research database held by me will be labeled by study IDs and not participants' names. Access to this database will be restricted to myself, my supervisory team (Dr N.Hanchard, Dr H. Handoll, Prof A. Batterham) and my advisor Dr Betthäuser.
h) Following a request from the Ethics commission of the Hamburg Medical Council, the data will be fully anonymised as soon as possible after being checked and cleared, i.e. prior to the statistical analysis.
i) The anonymised data will be archived on the secure servers at Teesside University for ten years.

All eligible patients/participants will be informed on all relevant aspects surrounding data collection, storage and confidentiality on the PIS (see Appendix 6.10).

12.3 Interventional Issues

I can foresee only minimal risks to the *interests* of participants, as there will be no deviation from standard clinical practice. Regarding the physiotherapy treatment of patients with PTT, there are no known serious risks or harms. A short-term increase in symptoms might occur as a temporary reaction to the physiotherapy treatment. Patients are routinely warned that this may occur and, should it do so, the treating physiotherapist will be the primary person to manage it. Participants will further be informed that they can at any time contact Dr Betthäuser, who will, as routine, monitor the whole treatment phase. Dr Betthäuser will also initiate and monitor provision of any adjunctive medical treatment (e.g. oral pain medication, local steroid injections to the shoulder), or any modifications of the treatment regime if considered necessary.

All participants will be informed (see PIS, Appendix 6.10) that, should any abusive or unprofessional behaviours or actions be disclosed and/or discovered, then confidentiality will be breached and Dr Betthäuser and the Director of Studies, Dr Hanchard will be informed.

In case of concerns/complaints about the conduct of the study, the details of a person who knows about the study but who is not involved in it (Dr Alasdair MacSween) will be provided.

12.4 Informed consent

Written informed consent will be obtained from all participants. See section 6 for details on the consent process. See Appendix 6.10 for the Patient Information Sheet (PIS) and Appendix 6.11 for the Consent Form.

14. Insurance liability issues

Insurance indemnity for this observational study was clarified following a request from Dr Alasdair MacSween in April 2011 and it was confirmed that the University's Public Liability and Professional Indemnity cover will be operative for this observational study. The relevant documents and correspondence have been forwarded to Dr Alasdair MacSween.

In addition, I will provide to the Chair of the School of Health & Social Care RG&E Committee written evidence of collaboration and indemnity for each potential collaborating physiotherapy practice before its actual involvement in the study (provided with ethics applications, see Appendix 6.6).

No payments will be made to participants.

15. References

Ainsworth, R. & Lewis, J. S. (2007). Exercise therapy for the conservative management of full thickness tears of the rotator cuff: a systematic review. *Br. J. Sports Med.*, 41(4), 200–10.

Beaudreuil, J.; Bardin, T.; Orcel, P. & Goutallier, D. (2007). Natural history or outcome with conservative treatment of degenerative rotator cuff tears. *Joint. Bone. Spine*, 74(6), 527–9.

Beaudreuil, J.; Dhénain, M.; Coudane, H. & Mlika-Cabanne, N. (2010). Clinical practice guidelines for the surgical management of rotator cuff tears in adults. *Orthop. Traumatol. Surg. Res.*, 96(2), 175–9.

Bouwmeester, W.; Zuithoff, N. P. a; Mallett, S.; Geerlings, M. I.; Vergouwe, Y.; Steyerberg, E. W.; Altman, D. G. & Moons, K. G. M. (2012). Reporting and methods in clinical prediction research: A systematic review. *PLoS Med.*, 9(5), 1–12.

Bundesdatenschutzgesetz. (2009). German Federal Data Protection Act. Available at: http://www.gesetze-im-internet.de/englisch_bdsg/ [Last accessed 15 May 2015].

Data Protection Act. (1998). Data Protection Act. Available from: http://www.legislation.gov.uk/ukpga/1998/29/contents [Last accessed 15 May 2015].

Finnan, R. P. & Crosby, L. a. (2010). Partial-thickness rotator cuff tears. *J. Shoulder Elbow Surg.*, 19(4), 609–16.

GCP-Verordnung. (2004). *Verordnung über die Anwendung der Guten Klinischen Praxis bei der Durchführung von klinischen Prüfungen mit Arzneimitteln zur Anwendung am Menschen*. Available at: http://www.gesetze-im-internet.de/bundesrecht/gcp-v/gesamt.pdf [Last accessed 15 May 2915].

Hamburgisches Datenschutzgesetz (1990). Hamburg Data Protection Act. Available at: http://www.landesrecht-hamburg.de/jportal/portal/page/bshaprod.psml?showdoccase=1&doc.id=jlr-DSGHA1990rahmen&st=lr [Last accessed 15 May 2015]

Huisstede, B. M. a; Koes, B. W.; Gebremariam, L.; Keijsers, E. & Verhaar, J. a N. (2011). Current evidence for effectiveness of interventions to treat rotator cuff tears. *Man. Ther.*, 16(3), 217–30.

Moons, K. G. M.; Royston, P.; Vergouwe, Y.; Grobbee, D. E. & Altman, D. G. (2009). Prognosis and prognostic research: what, why, and how? *BMJ*, 338, b375.; doi:10.1136/bmj.b375.

Peduzzi, P.; Concato, J.; Kemper, E.; Holford, T. R. & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.*, 49(12), 1373–9.

Piitulainen, K.; Ylinen, J.; Kautiainen, H. & Häkkinen, A. (2012). The relationship between functional disability and health-related quality of life in patients with a rotator cuff tear. *Disabil. Rehabil.*, 34(24), 2071–5.

Reilly, P.; Macleod, I.; Macfarlane, R.; Windley, J. & Emery, R. J. H. (2006). Dead men and radiologists don't lie: a review of cadaveric and radiological studies of rotator cuff tear

prevalence. *Ann. R. Coll. Surg. Engl.*, 88(2), 116–21.

Seida, J. C.; LeBlanc, C.; Schouten, J. R.; Mousavi, S. S.; Hartling, L.; Vandermeer, B.; Tjosvold, L. & Sheps, D. M. (2010). Systematic review: nonoperative and operative treatments for rotator cuff tears. *Ann. Intern. Med.*, 153(4), 246–55.

Smith, M. A. & Smith, W. T. (2010). Rotator cuff tears: an overview. *Orthop. Nurs.*, 29(5), 319–22; quiz 323–4.

University of Teesside. (2007). Guidelines for good conduct in research. Available at: https://www.tees.ac.uk/docs/docrepo/research/GuidelinesforGoodConductinResearch.pdf [Last accessed 15 May 2015].

Vittinghoff, E. & McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. *Am. J. Epidemiol.*, *165*(6), 710–8.

World Medical Association. (2008). *Declaration of Helsinki, Version 2008*. Available at: http://www.wma.net/en/30publications/10policies/b3/17c.pdf [Last accessed 15 May 2015].

Appendix 6.4
Deviations from the study protocol

No	Aspect	**Deviation,** with key reason and reference to relevant thesis sections (Chapters 6 and 7)
1	Study personnel: collaborating physiotherapist practices	The initial number of collaborating physiotherapy practices was expanded to improve recruitment. See section 6.6.3.2
2	Sample size consideration, analysis of primary outcome	The initial sample size consideration was based on the analysis of the WORC_change as a binary outcome based on the estimated MID derived from the sample data. Subsequently, to avoid the unnecessary loss of information that would result from dichotomisation of the outcome, I decided to analyse the WORC_change on a continuous scale. See section 6.6.14
3	Univariable analysis of pain catastrophizing (PCS)	I conducted a supplementary analysis to explore the contribution of this factor alone in predicting the worc change. This was justified by the observation that, alone among the factors, pain catastrophizing was part of the two “best” models and the further “plausible” alternative model. See section 6.7.7.4
4	Responder analysis	The decision to complement the MID analysis by the exploratory responder analysis was made after the protocol had been completed. The intention was to enhance the interpretation of the WORC_change scores based on the MID estimate. See Chapter 7 (sections 7.1, 7.3)

□□□□□□□□□

□□ □□ □ □□□□□□□□□[19]

TRIPOD Checklist: Prediction Model Development

Section/ Topic	Item	Checklist Item	*Key section in Ch. 6*
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	*Title, 6.6.9.1*
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	*n/a (thesis abstract only)*
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	*6.1*
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	*6.2*
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	*6.6.1*
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	*6.7.1*
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	*6.6.2-3*
	5b	Describe eligibility criteria for participants.	*6.6.4*
	5c	Give details of treatments received, if relevant.	*6.6.6*
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	*6.6.9.1*

[19] Retrieved from: https://www.tripod-statement.org/ [Last accessed 27 June 2016]

□

Section/ Topic	Item	Checklist Item	*Key section in Ch. 6*
	6b	Report any actions to blind assessment of the outcome to be predicted.	*6.6.9.3*
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	*6.6.7*
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	*n/a (prospective design)*
Sample size	8	Explain how the study size was arrived at.	*6.6.14*
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	*6.6.15, 6.7.2*
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	*6.6.17.1*
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	*6.6.17.5-9*
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	*6.6.17.7, 6.6.18, (6.6.17.9)*
Risk groups	11	Provide details on how risk groups were created, if done.	*n/a*
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	*6.7.1, 6.7.5*
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	*6.7.3 (charact.), 6.7.2 (missing data)*
Model development	14a	Specify the number of participants and outcome events in each analysis.	*6.7.7.1*
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	*n/a*
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	*6.7.7.2 (model statistics)*
	15b	Explain how to the use the prediction model.	*n/a: see 6.8.7*
Model performance	16	Report performance measures (with CIs) for the prediction model.	*6.7.7.2-3 (model statistics)*

Section/ Topic	Item	Checklist Item	*Key section in Ch. 6*
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	*6.8.4-8 (discussion)*
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	*6.8.9 and see above (18)*
Implications	20	Discuss the potential clinical use of the model and implications for future research.	*6.8.10*
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	*6.4 (protocol)*
Funding	22	Give the source of funding and the role of the funders for the present study.	*n/a (none)*

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

Appendix 6.6

Statement on cooperation and insurance liability form

C.Braun, Teesside University, May 2012_June 2014

Erklärung zu Kooperation und Versicherungsschutz Physiotherapie-Praxen
[Statement on cooperation and insurance liability of physiotherapy practices]

Studie: „Prognostische Faktoren für das Ergebnis konservativer Therapie für Patienten mit partiellen Rotatorenmanschettenrupturen“
[Study: “Predicting the outcome of conservative treatment for partial-thickness rotator cuff tears”]

Hiermit bestätige ich, dass Cordula Braun mich, stellvertretend für meine Praxis für Physiotherapie (Praxisnamen und Ort eintragen [insert name and place of practice])

_____

_____

über das o.g. Studienprojekt informiert hat, und mich zur Mitwirkung als Kooperationspraxis bei der Übernahme/ Durchführung der physiotherapeutischen Behandlung von Studienteilnehmern eingeladen hat. Ich habe mich zu dieser Kooperation bereit erklärt, die keinerlei gegenseitige Verpflichtungen beinhaltet.

[English translation: I herewith confirm that Cordula Braun informed me (on behalf of my practice for physiotherapy) about the above study project, and that she invited me to contribute as a cooperative practice by taking over/ carrying through the physiotherapeutic treatment of study participants. I consent to this cooperation that does not comprise any mutual obligations].

Ich bestätige, dass alle Physiotherapeuten meiner Praxis, die sich an der Behandlung der Studienpatienten beteiligen, einen ausreichenden Berufshaftpflicht-Versicherungsschutz haben.

[English translation: I confirm that all physiotherapists working in my practice who will be involved with the treatment of study participants, have sufficient professional insurance liability cover].

Im Fall der späteren Unterzeichnung: Versicherungsschutz bestand von Beginn der Kooperation an *[English translation: In case of delayed signature: I confirm that this insurance liability cover has been in place from the beginning of the collaboration].*

Name Unterzeichner (Praxisinhaber) [*Name signatory (Owner of practice*)]:

_____

Datum, Unterschrift *[Date, Signature]; ggf. Stempel [plus stamp if available)]*:

□

Appendix 6.7

Physical assessment of the shoulder - DVSE guideline

The table provides an overview of the guideline by Deutsche Vereinigung für Schulter- und Ellbogenchirurgie (DVSE) (2012)[20] on the assessment of the shoulder. The publication includes detailed descriptions of all tests, complemented by illustrative images. The components and tests should be chosen based on the individual patient's presentation and are not meant to constitute an exhaustive list.

Components	Recommended tests
1. General assessment/tests	
Assessment of shoulder mobility	"Neutral-Null" method, goniometry
Isometric tests	Abduction, external rotation, internal rotation
Muscle strength tests	Manual Muscle Functioning Test (MFP) 0-5
Capsular restrictions	Cyriax
Hyperlaxity	Local: anterior/posterior translation (anterior/posterior drawer test); sulcus sign, hyperabduction test (Gagey), Coudane-Walch test, supination-elbow extension test (SEET) General: Beighton Score/hypermobility score
2. Instability	
Ventral Instability	Apprehension test, relocation test, surprise test/release test, load-and-shift test;
Dorsal Instability	Load- and shift test/Norris test, Jerk test, Kim test
3. Impingement/rotator cuff lesions	
Impingement	Painful arc, Hawkins test, Neer test
Supraspinatus	Codman grip, Jobe test/empty can test/full can test, drop arm sign
Infraspinatus	Hornblower's sign, external rotation lag sign/dropping sign
Subscapularis	Belly press test, belly off sign, bear hug test, lift off test, internal rotation lag sign
Deltoideus	Deltoideus extension lag sign
4. Biceps tests	
Long head of biceps (LHB)/Sulcus/Pulley	Sulcus test (DePalma), Speed's test, O'Brien test/active compression

[20] See Chapter 6 reference list for the full reference.

Components	Recommended tests
Superior Labrum Anterior Posterior (SLAP) tests	Crank test/sign, O'Brien test/active compression test, supine flexion resistance test
5. Acromioclavicular (AC) joint	
	Finger sign, local pain on palpation, horizontal adduction test/cross body sign, LA (injection) test, piano key phenomenon, horizontal instability, upper painful arc
6. Scapular provocation tests	
	Scapular assistance test, lateral scapular slide test, scapular dyskinesis (Kibler), scapula alata ("pseudo-winging", scapular winging)
7. Thoracic outlet tests	
	Adson test, Eden test

Appendix 6.8
Initial assessment questionnaire - prognostic and baseline factors

Additional questions [part of initial assessment questionnaire package]

Additional questions - please respond to the following questions:

1. “What is the worst amount of pain that you have experienced within the past week?” Please respond by putting a slanted mark (“/“) on the below line, between “no pain” (left end) and “extreme pain” (right end).

2. “For how long have you been having your current shoulder complaints?” Please specify the duration in weeks.

_____ week/s

3. "Before you had your current shoulder problem, did a typical week include one or more of the following activities:

- Repetitive or prolonged use of the affected arm for strength effort (e.g. lifting, carrying or moving heavy loads, athletic sports, strength-demanding skilled manual work)

- Repetitive or prolonged use of the arm above shoulder height (e.g. overhead work, overhead sports, throwing sports, work as a hairdresser)?” (yes/no)Bitte antworten Sie durch Ankreuzen der für Sie zutreffenden Antwort („ja“ oder „nein“).

Please tick „yes“ if at least one of these activities applies to you.

Yes		No	

4. “Prior to the current episode, have you ever seen a medical doctor or therapist for pain in this shoulder?” Please respond by ticking the appropriate box („yes“ or „no“).

Ja		Nein	

5. “Are you a smoker? Please respond by ticking the appropriate box (“yes” or “no”). Tick “yes” if you regularly smoke at least once a week any amount of tobacco”

Ja		Nein	

(Question 6: see next page)

?

Additional questions [part of initial assessment questionnaire package]

6. What is your current work (occupational) status?
Please respond by ticking the appropriate box.

a) I am working full-time	
b) I am working part-time	
c) I am not working - retired	
d) I am not working – on sick leave	
e) I am not working – other reason	

Appendix 6.9
Global Perceived Change (GPC) scale

Global Perceived Change (GPC)[1]

We are interested to learn how you perceive the overall change of your shoulder problem since your first assessment with Dr. Betthäuser. Please make your rating by ticking (i.e. entering an "X" in) the box you consider appropriate:

"Since my first assessment with Dr Betthäuser, I rate my shoulder problem as…"

-3	-2	-1	0	1	2	3
Worse than ever	Much deteriorated	Slightly deteriorated	Unchanged	Slightly improved	Much improved	Completely recovered

[1] Part of follow-up assessment questionnaire package

Appendix 6.10

Patient information sheet (PIS)

Teesside University
Teesside University is sponsoring this project for the purposes of research governance

Schulter-Zentrum • Dr. A. Betthäuser • c/o Ev. Krankenhaus Alsterdorf • Elisabeth-Flügge-Str. 1 •
22337 Hamburg • Tel.: 460 735 50 • Email: info@schulter-zentrum.com
Kontakt C.Braun: Tel.: 0176 999 628 68 • Email: braun@hs21.de

Schulter-Zentrum

Consent to participate in a research study
(version 3_10/03/2012)

Study: "Predicting the outcome of conservative treatment with physiotherapy for shoulder pain in the presence of atraumatic partial-thickness tears of the rotator cuff"

German title: "Prognostische Faktoren für das Ergebnis konservativer Therapie mit Physiotherapie für Patienten mit Schulterschmerzen und nicht-traumatischen partiellen Rotatorenmanschettenrupturen"

I have received, read and understood the information sheet for the above study [version 2_06/28/2012]. I have been able to ask questions and have had these answered satisfactorily. I understand that taking part is voluntary and that I can withdraw at any time without giving a reason. I was given sufficient time to decide about my participation.
I give permission for Cordula Braun to access and use some information from my medical record held by Dr Betthäuser, and for her to receive some information from my treating physiotherapist. I give permission to Cordula Braun to contact Dr Betthäuser and my treating physiotherapist for clarification of questions if needed.
I give permission for Cordula Braun to write to me to remind me to make/remake a follow-up appointment with Dr Betthäuser or to ask me to complete the follow-up questionnaires for this episode of care.
If I ask for a copy of this consent form to keep, I will be given one.

Data protection:
The personal data, that is, after receipt of consent from the study participant, collected for the purpose of the study, specifically any medical results, underlies the obligation to secrecy and the data protection regulations.
It will be recorded in paper form as well as on storage media *in Dr.med. Andreas Betthäuser's practice (Schulter-Zentrum Hamburg),* and will be stored being pseudo-anonymised[1] (coded) for the duration of *the data collection period (anticipated 2-3 years, but not longer than 5 years).* For pseudo-anonymisation (coding), the name and other characteristics of identification (e.g. parts of the birth date) are replaced by e.g. multi-digit combinations of letters or numbers, also called code, in order to preclude or substantially hamper the identification of the study participants.
Access to the "key" that enables a personal allocation of the study participant is, besides the study supervisor Dr Andreas Betthäuser, restricted to the organising researcher Cordula Braun.
The use of the data by Dr Betthäuser and Cordula Braun will, during the data collection period, happen with the data pseudo-anonymised, and will during this period be restricted to the compilation and organisation of the data (data management, not analysis). In order to allow for academic support on data organisation by the supervisors at Teesside University, the data (but not the key) will concurrently be stored in a password-protected digital folder on a secure server at Teesside University, GB for the duration of the data collection period. Access to this folder will be restricted to Cordula Braun and the supervisors at Teesside University. The supervisors will at no time have access to the "key". Any further transfer of

the collected data within the scope of this study will only be done with the data being anonymised. The same accounts for the publication of the study results.
The study participants have the right to claim information about the personal data that has been collected from them. As the study is observational, participants will be informed of any personal clinical results in line with standard practice.
This study has been advised by the responsible ethics committees. The competent federal state authority may possibly be granted access to the study records.
As soon as the aim of the study allows for this, i.e. after completion of the data collection and before the analysis, the key will be deleted and the collected data thereby be anonymised[2].
In the case of withdrawal of consent, the data that has already been collected will likewise be deleted or anonymised[2], and are further used in this form.
A withdrawal of data that has already been anonymised is not possible.
1 pseudo-anonymisation is the replacement of the name and of other characteristicas of identification by a tag for the purpose of precluding the identification of the person concerned, or to substantially hamper it.
2 anonymisation is the change of personal data in a way that particulars about personal or factual conditions are no longer, or only through disproportional effort in time, costs or working power, attributable to a specific natural person (§ 3 subparagraph 6 Federal Data Protection Act).

Option: I agree that Cordula Braun may contact me again by letter until July 2017 to ask me whether I might be interested in taking part in a further study (please circle). YES / NO

I agree to take part in the above stated study.

Name (please print)	Date	Signature

Witness of signature:

Name (please print)	Date	Signature

Information provider:

Name (please print)	Date	Signature

5. What do I have to do if I want to take part?

Step 1:

Arrange your follow-up appointment with Dr Betthäuser for 3 months from your initial assessment.

Step 2:

Dr Betthaeuser will have provided you with a prescription for physiotherapy as part of a numbered 'study pack' which also contains this information sheet, a consent form, a list of the collaborating physiotherapy practices; and a report form for your physiotherapist to complete at the end of your treatment. Choose one of the listed physiotherapy practices and arrange your appointment with them.

Step 3:

When you go to see your physiotherapist for the first time, take the complete study pack with you. Your physiotherapist will ask you whether you have been invited to take part in the study, and if so, whether you have decided to do so or not. If you do not want to take part, the study will not be mentioned again. If you have decided to participate, you will be asked to sign the consent form. If you wish, you will be given a copy of the consent form to keep.

Your physiotherapist will keep your prescription and the report form. Your physiotherapist will also need to know your personal study ID, which is on the study pack as well as on the report form.

Step 4:

You receive your standard treatment. The type, content or amount of treatment will be in no way regulated for the purpose of this study. Any assessments and treatments will be delivered in accordance with the current German healthcare and insurance regulations. You will be advised on the number and type of follow-up prescriptions for physiotherapy by your treating physiotherapist, in agreement with Dr Betthäuser. You are free to choose to contact and see Dr Betthäuser at any time during the treatment phase if you feel in need of further advice or adjunctive medical treatment.

Step 5:

Remember your follow-up assessment with Dr Betthäuser. Your physiotherapist will remind you. If you have not had your follow-up appointment by around 3 months after consenting to take part, Cordula Braun will send you a written reminder, enclosing the standard follow-up questionnaires and a stamped, addressed envelope for return by post, in case you do not want to, or are unable to, make an appointment. Your follow-up assessment or return of the questionnaires by post will mark the end of your involvement in this study.

6. What are the possible disadvantages or advantages and benefits of taking part? There are no direct disadvantages or advantages for you if you take part in this study.

7. What will happen to the information collected about me? This is explained in the standard data protection statement of the Ethics Committee of the Medical Council, Hamburg (version 05/04/2011), in the box below:

Data protection:
The personal data, that is, after receipt of consent from the study participant, collected for the purpose of the study, specifically any medical results, underlies the obligation to secrecy and the data protection regulations.
It will be recorded in paper form as well as on storage media *in Dr.med. Andreas Betthäuser's practice (Schulter-Zentrum Hamburg),* and will be stored being pseudo-anonymised[1] (coded) for the duration of *the data collection period (anticipated 2-3 years, but not longer than 5 years).* For pseudo-anonymisation (coding), the name and other characteristics of identification (e.g. parts of the birth date) are replaced by e.g. multi-digit combinations of letters or numbers, also called code, in order to preclude or substantially hamper the identification of the study participants.
Access to the "key" that enables a personal allocation of the study participant is, besides the study supervisor Dr Andreas Betthäuser, restricted to the organising researcher Cordula Braun.
The use of the data by Dr Betthäuser and Cordula Braun will, during the data collection period, happen with the data pseudo-anonymised, and will during this period be restricted to the compilation and organisation of the data (data management, not analysis). In order to allow for academic support on data organisation by the supervisors at Teesside University, the data (but not the key) will concurrently be stored in a password-protected digital folder on a secure server at Teesside University, GB for the duration of the data collection period. Access to this folder will be restricted to Cordula Braun and the supervisors at Teesside University. The supervisors will at no time have access to the "key". Any further transfer of the collected data within the scope of this study will only be done with the data being anonymised. The same accounts for the publication of the study results.

The study participants have the right to claim information about the personal data that has been collected from them. As the study is observational, participants will be informed of any personal clinical results in line with standard practice.
This study has been advised by the responsible ethics committees. The competent federal state authority may possibly be granted access to the study records.
As soon as the aim of the study allows for this, i.e. after completion of the data collection and before the analysis, the key will be deleted and the collected data thereby be anonymised[2].
In the case of withdrawal of consent, the data that has already been collected will likewise be deleted or anonymised[2], and are further used in this form.
A withdrawal of data that has already been anonymised is not possible.
1 **pseudo-anonymisation** is the replacement of the name and of other characteristicas of identification by a tag for the purpose of precluding the identification of the person concerned, or to substantially hamper it.
2 anonymisation is the change of personal data in a way that particulars about personal or factual conditions are no longer, or only through disproportional effort in time, costs or working power, attributable to a specific natural person (§ 3 subparagraph 6 Federal Data Protection Act).

The completed report of this study will be sent to and reviewed by academic staff at Teesside University, UK. The findings may also be presented at conferences and printed in journals, and will form part of Cordula Braun's submitted PhD thesis. You will not be identified in any presentation or journal or in the thesis. If you would like us to we will send you information on the results of the study.

8. What happens if I have a complaint about the conduct of this study?
If you have any complaint about the conduct of this study, please contact either Cordula Braun or Dr Betthäuser. Should any abusive or unprofessional behaviours or actions towards you be disclosed and/or discovered then confidentiality will be breached and if deemed appropriate further action(s) may be taken.

9. Who is organising and funding this study? This study is a non-funded study, i.e. it is not financially supported by anyone. It is organised by Cordula Braun and supervised by academic staff from Health and Social Care Institute, Teesside University, UK.

10. Who can I contact with any queries? For any queries or further information please contact the principal researcher as the primary contact person:

Cordula Braun	**Phone**	**0176 - 999 628 68**
	Email	**C.Braun@tees.ac.uk**

You can also contact

Dr Andreas Betthäuser	**Phone**	**040 – 460 735 50**
	Email	**info@schulter-zentrum.de**

Thank you for reading this information sheet and for considering whether or not to take part in the study

Appendix 6.11
Consent form

Schulter-Zentrum • Dr. A. Betthäuser • c/o Ev. Krankenhaus Alsterdorf • Elisabeth-Flügge-Str. 1 •
22337 Hamburg • Tel.: 460 735 50 • Email: info@schulter-zentrum.com
Kontakt C.Braun: Tel.: 0176 999 628 68 • Email: braun@hs21.de

Consent to participate in a research study
(version 3_10/03/2012)

Study: "Predicting the outcome of conservative treatment with physiotherapy for shoulder pain in the presence of atraumatic partial-thickness tears of the rotator cuff"

German title: "Prognostische Faktoren für das Ergebnis konservativer Therapie mit Physiotherapie für Patienten mit Schulterschmerzen und nicht-traumatischen partiellen Rotatorenmanschettenrupturen"

I have received, read and understood the information sheet for the above study [version 2_06/28/2012]. I have been able to ask questions and have had these answered satisfactorily. I understand that taking part is voluntary and that I can withdraw at any time without giving a reason. I was given sufficient time to decide about my participation.
I give permission for Cordula Braun to access and use some information from my medical record held by Dr Betthäuser, and for her to receive some information from my treating physiotherapist. I give permission to Cordula Braun to contact Dr Betthäuser and my treating physiotherapist for clarification of questions if needed.
I give permission for Cordula Braun to write to me to remind me to make/remake a follow-up appointment with Dr Betthäuser or to ask me to complete the follow-up questionnaires for this episode of care.
If I ask for a copy of this consent form to keep, I will be given one.

Data protection:
The personal data, that is, after receipt of consent from the study participant, collected for the purpose of the study, specifically any medical results, underlies the obligation to secrecy and the data protection regulations.
It will be recorded in paper form as well as on storage media *in Dr.med. Andreas Betthäuser's practice (Schulter-Zentrum Hamburg),* and will be stored being pseudo-anonymised[1] (coded) for the duration of *the data collection period (anticipated 2-3 years, but not longer than 5 years).* For pseudo-anonymisation (coding), the name and other characteristics of identification (e.g. parts of the birth date) are replaced by e.g. multi-digit combinations of letters or numbers, also called code, in order to preclude or substantially hamper the identification of the study participants.
Access to the "key" that enables a personal allocation of the study participant is, besides the study supervisor Dr Andreas Betthäuser, restricted to the organising researcher Cordula Braun.
The use of the data by Dr Betthäuser and Cordula Braun will, during the data collection period, happen with the data pseudo-anonymised, and will during this period be restricted to the compilation and organisation of the data (data management, not analysis). In order to allow for academic support on data organisation by the supervisors at Teesside University, the data (but not the key) will concurrently be stored in a password-protected digital folder on a secure server at Teesside University, GB for the duration of the data collection period. Access to this folder will be restricted to Cordula Braun and the supervisors at Teesside University. The supervisors will at no time have access to the "key". Any further transfer of

the collected data within the scope of this study will only be done with the data being anonymised. The same accounts for the publication of the study results.
The study participants have the right to claim information about the personal data that has been collected from them. As the study is observational, participants will be informed of any personal clinical results in line with standard practice.
This study has been advised by the responsible ethics committees. The competent federal state authority may possibly be granted access to the study records.
As soon as the aim of the study allows for this, i.e. after completion of the data collection and before the analysis, the key will be deleted and the collected data thereby be anonymised[2].
In the case of withdrawal of consent, the data that has already been collected will likewise be deleted or anonymised[2], and are further used in this form.
A withdrawal of data that has already been anonymised is not possible.
1 pseudo-anonymisation is the replacement of the name and of other characteristicas of identification by a tag for the purpose of precluding the identification of the person concerned, or to substantially hamper it.
2 anonymisation is the change of personal data in a way that particulars about personal or factual conditions are no longer, or only through disproportional effort in time, costs or working power, attributable to a specific natural person (§ 3 subparagraph 6 Federal Data Protection Act).

Option: I agree that Cordula Braun may contact me again by letter until July 2017 to ask me whether I might be interested in taking part in a further study (please circle). YES / NO

I agree to take part in the above stated study.

Name (please print)	Date	Signature

Witness of signature:

Name (please print)	Date	Signature

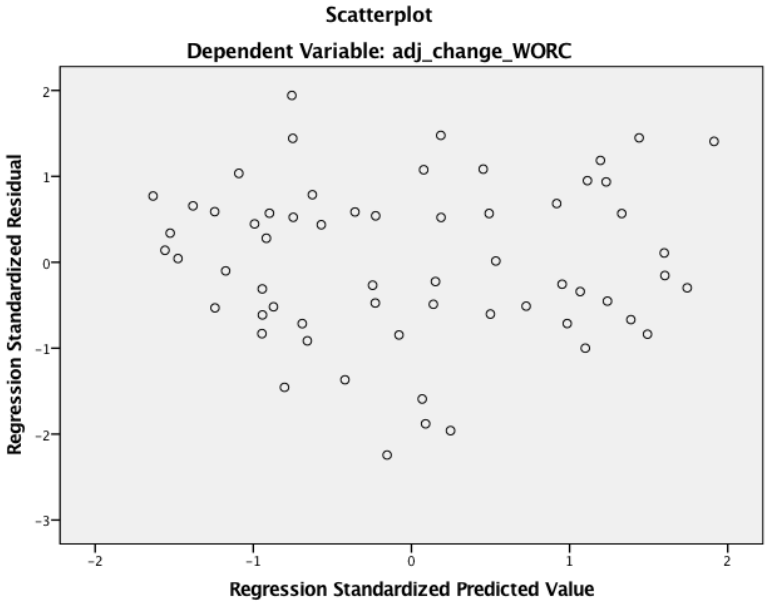
Information provider:

Name (please print)	Date	Signature

Appendix 6.12
Residual plots of prognostic models

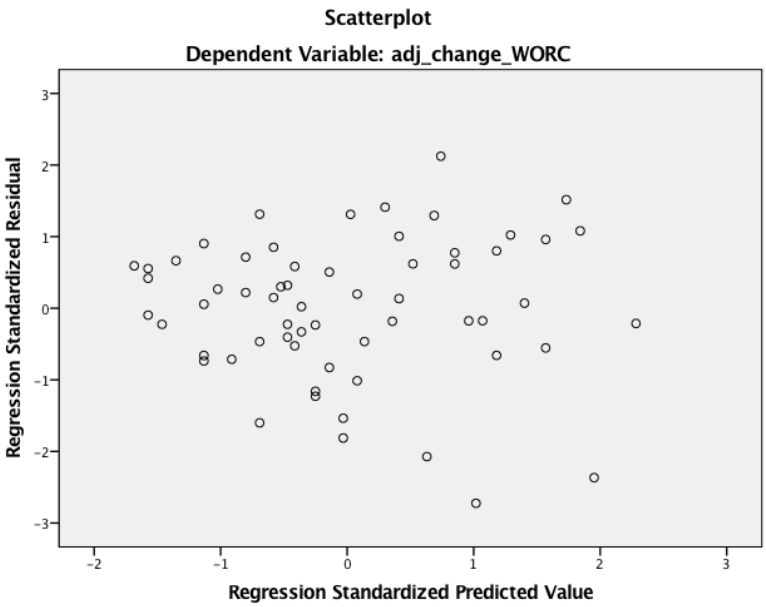
Model 1 (nine factors):

Age + sex + physical demands + disability (and QoL; WORC) + pain + history of shoulder pain + symptom duration + smoking + pain catastrophizing (PCS)



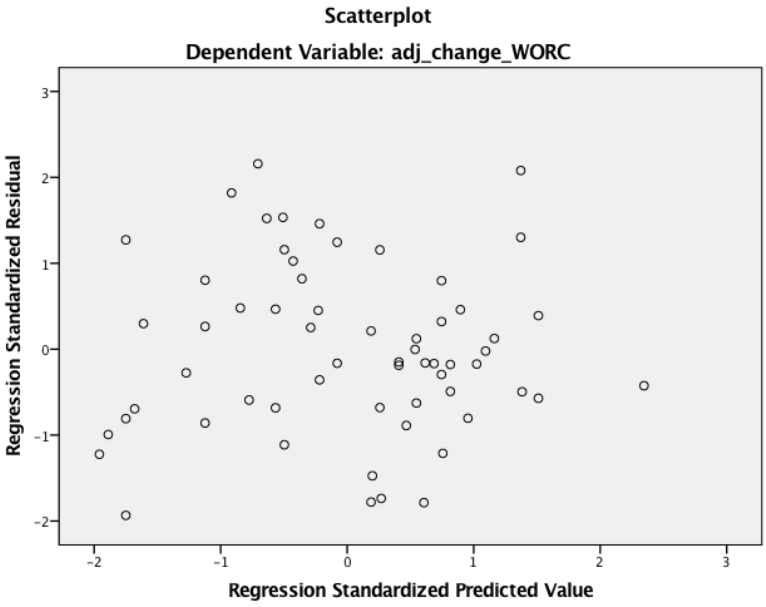
Model 2 (two factors):

Smoking + pain catastrophizing (PCS)



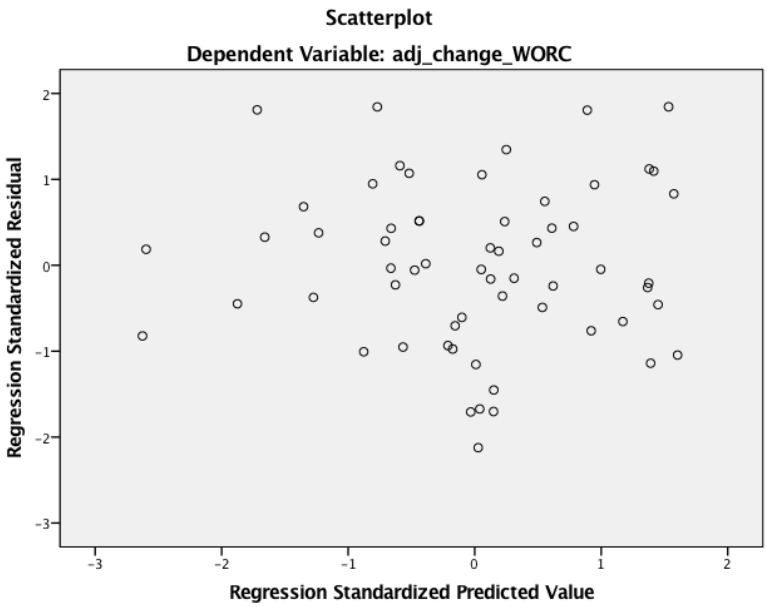
Model 3 (two factors):

Age + sex



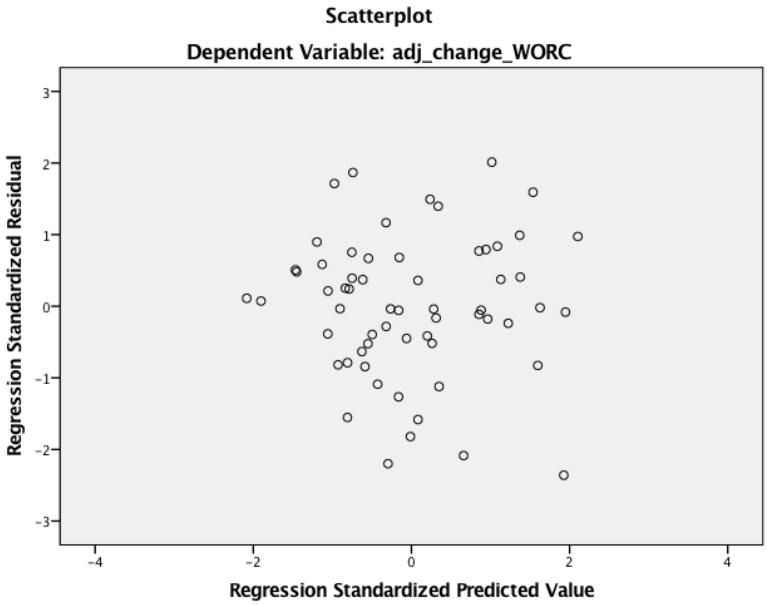
Model 4 (seven factors):

Age + sex + physical demands + pain + history of shoulder pain + symptom duration + smoking


Scatterplot
Dependent Variable: adj_change_WORC
Regression Standardized Residual
Regression Standardized Predicted Value
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2

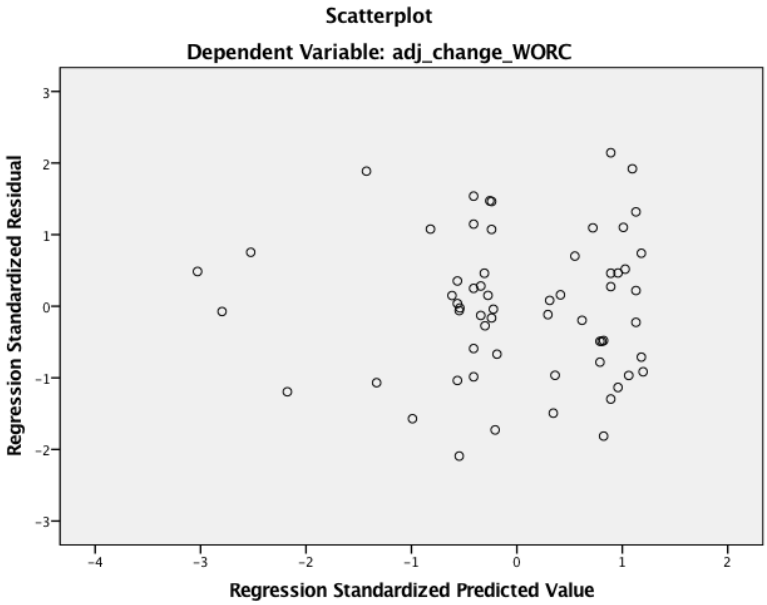
Model 5 (two factors):

Disability (and HrQoL; WORC) + pain catastrophizing (PCS)



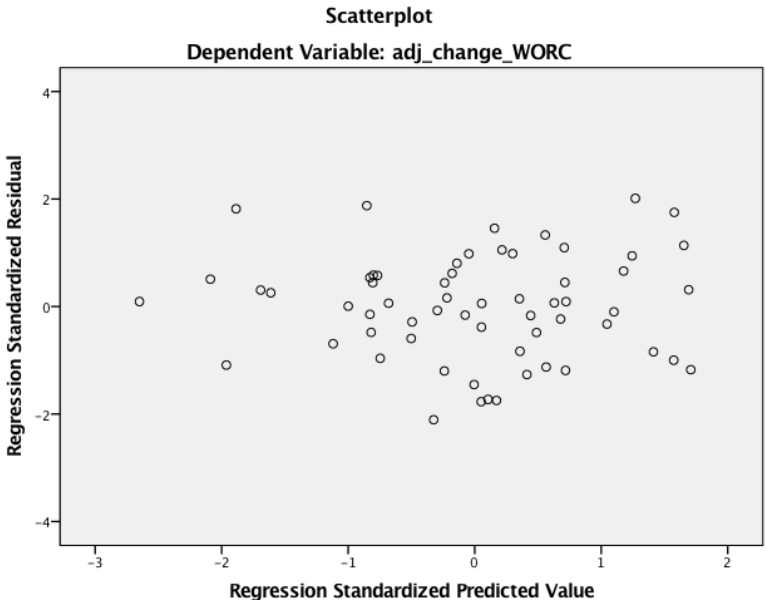
Model 7 (two factors):

History of shoulder pain + symptom duration



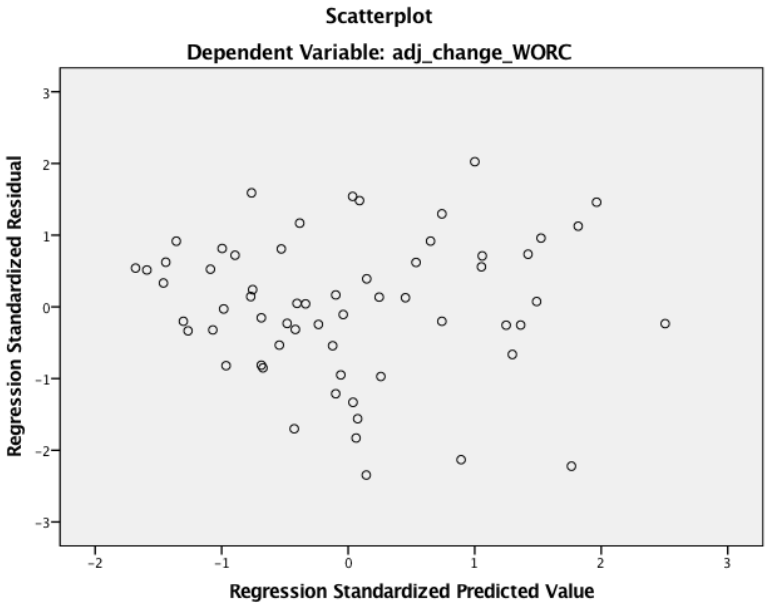
Model 8 (three factors)

Pain + history of shoulder pain + symptom duration

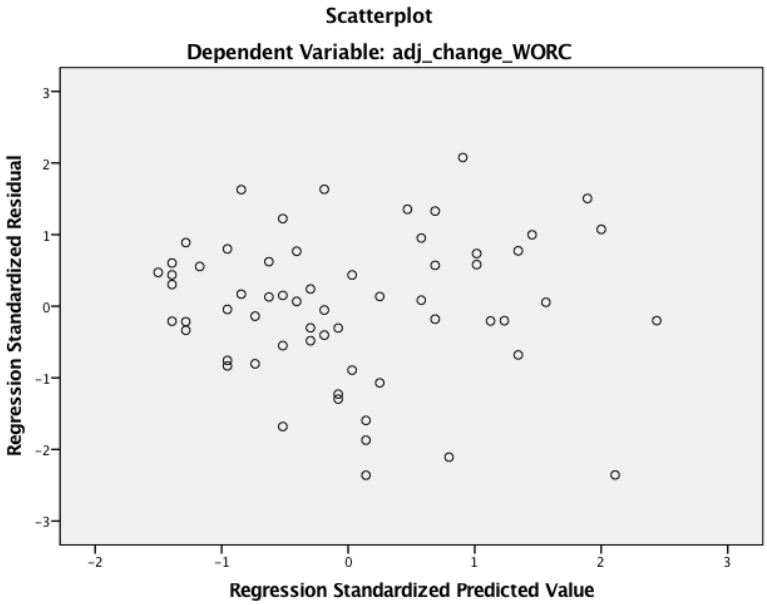


Model 9 (two factors)

Pain + pain catastrophizing (PCS)



Complementary pain catastrophizing (PCS) model:



Appendix 6.13

PROBAST version 20/07/2015[21]

Early version • Not intended for use in a review

PROBAST

(Prediction model study Risk Of Bias Assessment Tool)

Reference to published PROBAST papers…

What does PROBAST assess?

PROBAST assesses both the *risk of bias* and *applicability* of a study that evaluates (develops and/or validates) a multivariable diagnostic or prognostic prediction model. It is designed to assess primary studies included in a systematic review.

Bias occurs if systematic flaws or limitations in the design, conduct or analysis of a primary study distort the results. *Risk of bias* refers to the likelihood that a prediction model leads to distorted predictive performance for its intended use and targeted individuals. The predictive performance is typically evaluated using calibration, discrimination, and (re)classification. *Applicability* refers to the extent to which the prediction model from the primary study matches your systematic review question, for example in terms of the population or outcomes of interest.

A primary study may include the development and/or validation of more than one prediction model. A PROBAST assessment should be completed for each distinct model that is developed or validated for making individualised predictions. Only assess prediction models that are of interest to your systematic review.

PROBAST is not designed for all multivariable diagnostic or prognostic studies. For example, studies using multivariable models to identify predictors associated with an outcome but not attempting to develop a model for making individualised predictions are not covered by PROBAST.

PROBAST includes five steps.

Step	Task	When to complete
1	Specify your systematic review question	Once per systematic review
2	Classify the type of prediction model evaluation	Once for each model of interest in each publication being assessed, for each relevant outcome
3	Assess risk of bias and applicability	Once for each evaluation (development and/or validation) of each distinct model
4	Overall judgement	Once for each evaluation (development and/or validation) of each distinct model
5	Usability of the model	Once for each distinct model

If this is your first time using PROBAST, we recommend using the PROBAST-STARTER template which includes additional notes.

[21] The inclusion of this PROBAST version in the thesis appendices was approved by Dr Robert Wolff (personal communication, 17/12/2015).

Step 1: Specify your systematic review question

State your systematic review question to facilitate the assessment of the applicability of the evaluated models to your question. *The following table should be completed once per systematic review.*

Specify your systematic review question
Participants (e.g. setting, main inclusion criteria, prior treatments):
Outcome(s) to be predicted:
Intended use of the model(s): □ Diagnosis □ Prognosis *When will the model(s) be used, e.g. at presentation with signs/ symptoms, staging severity of disease, pre-operatively?*

Step 2: Classify the type of prediction model evaluation

Different signalling questions apply for different types of prediction model evaluation. Use the following table to classify the evaluation as model development, model validation or both. If the evaluation does not fit one of these classifications then PROBAST should not be used.

This table should be completed once for each model of interest in each publication being assessed, for each relevant outcome in your review.

Publication reference	
Model of interest	
Outcome of interest	

Classify the evaluation based on its aim		
Type of model evaluation	**Tick as appropriate**	**PROBAST classification**
Prediction model development without testing its predictive performance in other individuals, i.e. no external validation. Model development should ideally include internal validation, such as bootstrapping or cross-validation.		Development (Dev) only
Prediction model development as well as testing of predictive performance in other individuals (external validation), both in the same publication.		Development (Dev) and external validation (Val)
Testing the predictive performance of a previously developed prediction model in other individuals (external validation).		External validation (Val) only

Step 3: Assess risk of bias and applicability

PROBAST is structured as five key domains. Each domain is judged for risk of bias (low, high or unclear) and includes signalling questions to help make judgements. Signalling questions are rated as yes (Y), probably yes (PY), probably no (PN), no (N) or no information (NI). All signalling questions are phrased so that "yes" indicates absence of bias. Any signalling question rated as "no" or "probably no" flags the potential for bias; you will need to use your judgement to determine whether the domain should be rated as "high", "low" or "unclear" risk of bias. The guidance document contains further instructions and examples on rating signalling questions and risk of bias for each domain.

The first three domains are also rated for concerns regarding applicability (low/ high/ unclear) to your review question defined above.

Complete all domains separately for each evaluation of a distinct model. Shaded boxes indicate where signalling questions do not apply and should not be answered.

DOMAIN 1: Participant selection			
A. Risk of Bias			
Describe the sources of data and criteria for participant selection:			
		Dev	Val
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?			
2. Were all inclusions and exclusions of participants appropriate?			
3. Were participants enrolled at a similar state of health, *or* were predictors considered to account for differences?			
Risk of bias introduced by selection of participants	**RISK:** *(low/ high/ unclear)*		
Justification of bias rating:			
B. Applicability			
Describe included participants, setting and dates:			
Concern that the included participants and setting do not match the review question	**CONCERN:** *(low/ high/ unclear)*		
Justification of applicability rating:			

DOMAIN 2: Predictors			
A. Risk of Bias			
List and describe predictors assessed, e.g. definition, timing and knowledge of other predictors:			
		Dev	Val
1. Were predictors defined and assessed in a similar way for all participants?			
2. Were predictor assessments made without knowledge of outcome data?			
3. Are all predictors available at the time the model is intended to be used?			
4. Were all relevant predictors analysed?			
Risk of bias introduced by predictors or their assessment	**RISK:** *(low/ high/ unclear)*		
Justification of bias rating:			
B. Applicability			
Concern that the definition, assessment or timing of assessment of predictors in the model do not match the review question	**CONCERN:** *(low/ high/ unclear)*		
Justification of applicability rating:			

DOMAIN 3: Outcome			
A. Risk of Bias			
Describe the outcome and how it was defined and determined:			
		Dev	Val
1. Was a pre-specified outcome definition used?			
2. Were predictors excluded from the outcome definition?			
3. Was the outcome defined and determined in a similar way for all participants?			
4. Was the outcome determined without knowledge of predictor information?			
Risk of bias introduced by the outcome or its determination	**RISK:** *(low/ high/ unclear)*		
Justification of bias rating:			
B. Applicability			
At what time point was the outcome determined: *If a composite outcome was used, describe the relative frequency/distribution of each contributing outcome:*			
Concern that the outcome, its definition, timing or determination do not match the review question	**CONCERN:** *(low/ high/ unclear)*		
Justification of applicability rating:			

DOMAIN 4: Sample size and participant flow			
Risk of Bias			
Describe numbers of participants, outcome events and events per predictor: *Describe the time interval between predictor assessment and outcome determination:* *Describe any participants who were excluded from the model:* *Describe missing data on predictors and outcomes as well as methods used for missing data:*			
		Dev	Val
1. Were there a reasonable number of outcome events?			
2. Was the time interval between predictor assessment and outcome determination appropriate?			
3. Were all enrolled participants included in the analysis?			
4. Were participants with missing data handled appropriately?			
Risk of bias introduced by sample size or participant flow	**RISK:** *(low/ high/ unclear)*		
Justification of bias rating:			

DOMAIN 5: Analysis			
Risk of Bias			
Describe how the model was developed (predictor selection, fitting and optimism, risk groups, model performance): *Describe whether and how the model was validated, either internally (e.g. bootstrapping, cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants):* *Describe the performance measures of the model, e.g. calibration, discrimination, (re)classification, net benefit:*			
		Dev	Val
1. Were non-binary predictors handled appropriately?			
2. Was selection of predictors based on univariable analysis avoided?			
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?			
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?			
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?			
6. For the model or any simplified score, were relevant performance measures evaluated, e.g. calibration, discrimination, (re)classification and net benefit?			
7. Was the model recalibrated or was it likely (based on the evidence presented, e.g. calibration plot) that recalibration was not needed?			
Risk of bias introduced by the analysis	**RISK:** *(low/ high/ unclear)*		
Justification of bias rating:			

Step 4: Overall judgement

Use the following tables to reach overall judgements about risk of bias and applicability of the prediction model evaluation (development and/ or validation) across all assessed domains.
Complete for each evaluation of a distinct model.

Reaching an overall judgement about risk of bias of the prediction model evaluation	
Low risk of bias	If all domains were rated low risk of bias. If a prediction model was developed without any external validation, and it was rated as low risk of bias for all domains, consider downgrading to **high risk of bias**. Such model can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation.
High risk of bias	If at least one domain is judged to be at **high risk of bias**.
Unclear risk of bias	If an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Reaching an overall judgement about applicability of the prediction model evaluation	
Low concerns regarding applicability	If low concerns regarding applicability for all domains, the prediction model evaluation is judged to have **low concerns regarding applicability**.
High concerns regarding applicability	If high concerns regarding applicability for at least one domain, the prediction model evaluation is judged to have **high concerns regarding applicability**.
Unclear concerns regarding applicability	If unclear concerns (but no "high concern") regarding applicability for at least one domain, the prediction model evaluation is judged to have **unclear concerns regarding applicability** overall.

Overall judgement about risk of bias and applicability of the prediction model evaluation		
Overall judgement of risk of bias	**RISK:** *(low/ high/ unclear)*	
Justification of bias rating:		
Overall judgement of applicability	**CONCERN:** *(low/ high/ unclear)*	
Justification of applicability rating:		

Step 5: Usability of the model

The following question assesses whether the model was presented in enough detail to be usable in the targeted individuals and context. Note that this is different from the applicability assessment above, which refers to the extent to which the prediction model evaluation matches your review question.
Complete for each evaluation of a distinct model or simplified score.

Assess the usability of the model		
Is the model presented with sufficient detail to be used in the intended context and target population?	**RATING:** *(yes/ no)*	

Appendix 6.14
Detailed results of PROBAST assessment

See Appendix 3.8 for the PROBAST coding manual

Abbreviations and symbols used in the table:

Y = yes; PY = probably yes; N = no; NI = no information; N/A = not applicable; UNANSW = unanswerable (see coding manual); = low risk/concerns; = high risk/concerns (usability: not usable); ? = unclear risk/concerns

PROBAST items & judgements	Ratings (see above for explanations)
Domain 1: Participant selection	
1. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y
2. Were all inclusions and exclusions of participants appropriate?	PY*
3. Were participants enrolled at a similar state of health, or were predictors considered to account for differences?	Y
Risk of bias judgement	
Justification of bias rating: *The PY acknowledges the known limitations to the sensitivity/specificity of diagnosing PTT by ultrasonography	
Applicability judgement	
Justification of applicability rating: --	
Domain 2: Predictors	
1. Were predictors defined and assessed in a similar way for all participants?	Y
2. Were predictor assessments made without knowledge of outcome data?	Y
3. Are all predictors available at the time the model is intended to be used?	Y
4. Were all relevant predictors analysed?	UNANSW.
Risk of bias judgement	
Justification of bias rating: --	
Applicability judgement	
Justification of applicability rating: --	

PROBAST items & judgements	Ratings (see above for explanations)
Domain 3: Outcome	
1. Was a pre-specified outcome definition used?	Y
2. Were predictors excluded from the outcome definition?	Y†
3. Was the outcome defined and determined in a similar way for all participants?	Y
4. Was the outcome determined without knowledge of predictor information?	PY‡
Risk of bias judgement	□
Justification of bias rating: †This was rated as yes, because the WORC scores were adjusted for regression to the mean (RTM) (see body of thesis and PROBAST coding manual) ‡The PY acknowledges the fact that practical blinding was assumed (see body of thesis and PROBAST coding manual)	
Applicability judgement	□
Justification of applicability rating: --	
Domain 4: Sample size and participant flow	
1. Were there a reasonable number of outcome events?	Y
2. Was the time interval between predictor assessment and outcome determination appropriate?	Y
3. Were all enrolled participants included in the analysis?	Y
4. Were participants with missing data handled appropriately?	Y
Risk of bias judgement	□
Justification of bias rating: --	
Domain 5: Analysis	
1. Were non-binary predictors handled appropriately?	Y
2. Was selection of predictors based on univariable analysis avoided?	Y
3. Was model overfitting (optimism in model performance) accounted for, e.g. using bootstrapping or shrinkage techniques?	N
4. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	N/A
5. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	N/A
6. For the model or any simplified score, were relevant performance measures evaluated, e.g. calibration, discrimination, (re-) classification and net benefit?	N

□

PROBAST items & judgements	Ratings (see above for explanations)
7. Was the model recalibrated or was it likely (based on the evidence presented, e.g. calibration plot) that recalibration was not needed?	
Risk of bias judgement	□
Justification of bias rating: --	
Overall judgments	
Risk of bias	□
Applicability	□
Usability	□

Key: Y = yes; PY = probably yes; N = no; NI = no information; N/A = not applicable; UNANSW = unanswerable (see coding manual); □ = low risk/concerns; □ = high risk/concerns (usability: not usable); ? *= unclear risk/concerns*

Appendix 7.1
Screening of studies to inform the typical error of the WORC

Studies are listed in alphabetical order. The references are provided in the chapter reference list.

No	Study ID*	Result of screening: retained for further consideration ( ) or exclude ( )	Main reason for exclusion
1	Baskurt 2011		There was no untreated cohort.
2	Bernhardsson 2011		There was no untreated cohort.
3	de Witte 2012		
4	Dilek 2016		There was no untreated cohort.
5	Ekeberg 2010		There was no untreated cohort.
6	Ekeberg 2008		
7	Haik 2014		There was no untreated cohort.
8	Holtby 2005		There was no untreated cohort.
9	Kirkley 2003		
10	Kuhn 2013		There was no untreated cohort.
11	Lopes 2008		
12	Martins 2012		There was no untreated cohort.
13	Provencher 2012		WORC was assessed only once.
14	Razmjou 2008		WORC was assessed only once.
15	Razmjou 2006		There was no untreated cohort.
16	Subasi 2012		There was no untreated cohort.
17	Wessel 2005		WORC was assessed only once.
18	Wiertsema 2013		

** First author, year*

Appendix 7.2
Probabilities and confidence intervals (CIs) of individual WORC_change values

Values are ordered by magnitude (from the highest to the lowest $WORC\_change_{ADJ}$ value), and are consecutively numbered (from 1 to 65).

No	$WORC\_Change_{ADJ}$ value	Probability descriptor*	Probability	95% CI	
				lower limit	upper limit
1	-1102	alm_cert	1.0	-1510	-695
2	-964	alm_cert	1.0	-1372	-556
3	-948	alm_cert	1.0	-1355	-540
4	-946	alm_cert	1.0	-1354	-539
5	-872	alm_cert	1.0	-1280	-465
6	-861	alm_cert	1.0	-1269	-453
7	-841	alm_cert	1.0	-1249	-434
8	-792	alm_cert	0.99	-1200	-385
9	-770	very_lik	0.99	-1178	-363
10	-748	very_lik	0.98	-1155	-340
11	-724	very_lik	0.98	-1131	-316
12	-713	very_lik	0.98	-1121	-305
13	-682	very_lik	0.97	-1090	-275
14	-652	very_lik	0.96	-1060	-244
15	-629	likely	0.94	-1036	-221
16	-607	likely	0.93	-1015	-200
17	-602	likely	0.93	-1009	-194
18	-591	likely	0.92	-999	-183
19	-561	likely	0.90	-968	-153
20	-523	likely	0.86	-931	-115
21	-504	likely	0.84	-911	-96
22	-503	likely	0.84	-911	-95
23	-499	likely	0.83	-906	-91
24	-479	likely	0.81	-887	-71
25	-440	likely	0.75	-847	-32
26	-432	possibly	0.74	-840	-25

No	WORC_ $Change_{ADJ}$ value	Probability descriptor*	Probability	95% CI	
				lower limit	upper limit
27	-419	possibly	0.72	-826	-11
28	-413	possibly	0.71	-820	-5
29	-406	possibly	0.70	-814	2
30	-400	possibly	0.69	-808	8
31	-400	possibly	0.69	-807	8
32	-398	possibly	0.68	-805	10
33	-386	possibly	0.66	-794	21
34	-351	possibly	0.60	-759	57
35	-345	possibly	0.59	-753	62
36	-335	possibly	0.57	-743	73
37	-333	possibly	0.56	-740	75
38	-308	possibly	0.52	-716	99
39	-302	possibly	0.50	-710	106
40	-294	possibly	0.49	-702	113
41	-280	possibly	0.46	-688	128
42	-250	possibly	0.40	-658	157
43	-242	possibly	0.39	-650	165
44	-233	possibly	0.37	-640	175
45	-232	possibly	0.37	-640	175
46	-179	possibly	0.28	-587	228
47	-175	possibly	0.27	-583	232
48	-154	unlik	0.24	-562	254
49	-152	unlik	0.24	-560	255
50	-114	unlik	0.18	-521	294
51	-104	unlik	0.17	-512	303
52	-73	unlik	0.14	-480	335
53	-48	unlik	0.11	-456	360
54	-24	unlik_harm†	0.09	-431	384
55	-6	unlik_harm†	0.08	-414	401
56	26	unlik_harm†	0.06	-381	434
57	42	unlik_harm†	0.05	-366	449
58	108	unlik_harm‡	0.02	-300	515
59	110	unlik_harm‡	0.02	-298	518

No	WORC_ Change$_{ADJ}$ value	Probability descriptor*	Probability	95% CI	
				lower limit	upper limit
60	120	unlik_harm‡	0.02	-288	527
61	125	unlik_harm‡	0.02	-283	533
62	197	poss_harm‡‡	0.01	-210	605
63	288	poss_harm‡‡	0.00	-120	695
64	321	poss_harm‡‡	0.00	-87	728
65	387	poss_harm‡‡	0.00	-21	794

**Key: alm_cert = almost certainly beneficial; very_lik = verly likely beneficial; likely = likely beneficial; possibly = possibly beneficial; unlik = unlikely beneficial; unlik_harm = unlikely harmful; poss_harm = possibly harmful;*
†the verbal description related to a beneficial outcome for this case was "unlikely beneficial"; ‡the verbal description related to a beneficial outcome for this case was "very unlikely beneficial"; ‡‡the verbal description related to a beneficial outcome for this case was „almost certainly not beneficial".